Évaluer l'apprentissage par ordinateur : Tâches et rétroactions

Philippe Dessus

Laboratoire des sciences de l'éducation (EA-602), UPMF Grenoble-2 et Espé-UJF Grenoble-1

Maison page : http://webcom.upmf-grenoble.fr/sciedu/pdessus/

Courriel : `Philippe.Dessus@upmf-grenoble.fr`

Version 2.3 au 6 mars 2015

RESUME

Ce cours de Master 2 R sciences de l'éducation à distance a pour but de présenter des éléments théoriques et pratiques liés à la constitution de questionnaires à choix multiple (QCM) pour évaluer l'apprentissage. Il est composé de trois parties.
La première présente les éléments théoriques (notion de tâche, de feedback, panoplie de moyens d'évaluer l'apprentissage). La deuxième partie se centre sur les questionnaires à choix multiple et donne quelques indications pour utiliser un logiciel de conception de QCM. La troisième partie décrit les objectifs et l'organisation du cours. Elle intéressera donc principalement les étudiants suivant ce cours.

Note : Nous nous adressons tout au long de ce cours à un étudiant plutôt qu'à une étudiante. Que ces dernières soient assurées que c'est uniquement par souci de lisibilité et ne correspond à aucune discrimination à leur encontre.

Ce cours peut être cité de la manière suivante : Dessus, P. (2015). Evaluer l'apprentissage par ordinateur : tâches et rétroactions. Rouen : Cours de séminaire TICE du MARDIF, CNED.

Table des matières

Préambule

Selon leurs intérêts, les personnes lectrices de ce cours pourront l'aborder de différentes manières. Si elles sont étudiantes au MARDIF et suivent le séminaire TICE d'avril, il est préférable qu'elles passent directement à la troisième partie du document, qui contient la description des tâches liées à ce séminaire.

Ce cours en est à sa 5e édition (avec des actualisations mineures, essentiellement sur la § 7). La 4e édition avait un ajout principal dans la § 9e (la présentation d'*Oppia*). La 3e édition a été légèrement augmentée (le calcul de l'indice *B* et ajout de quelques ressources). La 2e édition était entièrement révisée (une douzaine de pages supplémentaires) et en partie refondue (3e partie), avec l'ajout de la section 10.

Tous les autres publics (enseignants, étudiants) doivent savoir que les deux autres parties de ce cours peuvent se lire indépendamment l'une de l'autre. Les personnes intéressées par les aspects pratiques de la construction de QCM passeront directement à la 2e partie. Les autres liront le document au fil des pages…

Bonne lecture et, surtout, bon travail !

Philippe Dessus, 6 mars 2015

Remerciements

L'auteur remercie vivement Christophe Charroud, Olivier Cosnefroy, Cédric D'Ham, Hussein Atta Salem et Emmanuelle Villiot-Leclercq pour leurs commentaires d'une version antérieure de ce cours. Olivier Cosnefroy et Thierry Rocher, par leurs commentaires vigilants, ont fait en sorte que la section 10 soit plus claire et moins ambiguë. Jérôme Panchout, enfin, par son aide efficace et proactive, a permis de rendre l'organisation des premiers de ces cours MARDIF plus simple et claire.

Selon la formule, les erreurs qui peuvent subsister dans le cours et son organisation (choisir la bonne réponse) :

a : ~~sont entièrement de la faute des collègues précédemment cités~~

b : sont de la pleine responsabilité de l'auteur !

Je crois que les questions à choix multiple :

a. Sont une méthode d'évaluation désuète.

b. Mesurent seulement des rappels de faits.

c. Devraient n'avoir aucun rôle dans l'enseignement des sciences

d. Toutes les réponses ci-dessus.

e. Aucune des réponses ci-dessus

Sadler (1998, p. 249)

1. Introduction

Tout étudiant qui connaît sur le bout du doigt son système éducatif sait que, s'il est facile d'éviter un mauvais cours, il est en revanche plus difficile d'éviter une mauvaise évaluation (Kalz, 2009). Dans le domaine de l'enseignement à distance, on évalue (ou simplement vérifie) les cours par le biais de questionnaires dont les retours sont plus ou moins immédiats. C'est un moyen courant et maintenant bien géré par de nombreuses plates-formes d'enseignement à distance (p. ex., *Moodle*). Il existe même des standards, permettant de transférer aisément d'une plate-forme à l'autre les questionnaires réalisés (comme SCORM : *Shareable content object reference model*, IMS QTI, *Question and test interoperability*, ou encore AICC, *Aviation industry CBT committee*).

Tout enseignant passe beaucoup de temps à essayer de comprendre ce que ses élèves arrivent à comprendre à propos des cours qu'il conçoit et réalise. Une partie de cette activité se passe en direct, *via* l'observation des élèves, leur interrogation, la correction d'exercices. Cette charge de travail est très importante, et l'enseignant peut penser à recourir à des outils informatisés qui l'aident à fournir de telles informations. Bien évidemment, ce type d'activités existait bien avant l'invention de l'ordinateur. On peut penser aux « boîtes enseignantes » de Freinet, où un élève pouvait sélectionner un ensemble de questions sur un domaine donné, l'insérer dans une boîte laissant successivement apparaître dans une fenêtre des questions et leurs réponses, le laissant ainsi s'auto-évaluer. De nombreux didacticiels informatisés qui sont ensuite apparus ne réalisent souvent pas autre chose (par exemple, lorsque les élèves utilisent un logiciel de traitement de textes et qu'ils sollicitent des suggestions lorsqu'un mot est indiqué comme mal orthographié), voir Section 2.

Les questionnaires à choix multiple, ou encore les simples questions-réponses sont un moyen classique de vérifier si un apprenant a compris. Ce type de questionnaires existe depuis de nombreuses années et a déjà fait l'objet de publications nombreuses (voir Thomas M. Haladyna, 2004 pour une synthèse). Pour autant, il est assez curieux de noter qu'il y a un engouement récent, dû à notre avis à deux raisons : – leur utilisation intensive dans l'enseignement à distance et plus, récemment, avec les boîtiers de vote ; – leurs effets sur l'apprentissage, qui est encore mal connu, et encore plus mal expliqué.

Ce cours permet d'étudier de plus près trois événements importants et leurs effets respectifs : l'apprentissage proprement dit, le rappel d'informations (dû, par exemple, à un questionnaire) et les rétroactions (de l'enseignant, d'une machine). Le sens commun signale que l'apprentissage a *uniquement* lieu pendant la phase d'exposition à l'information. Des recherches récentes (p. ex., Roediger & Butler, 2011) montrent que l'apprentissage de notions est d'autant meilleur qu'il est suivi de feedback, non pas immédiats à la question, mais reportés quelques minutes après (feedback différé). Ce cours s'intéresse donc à ce que Mayer et Colvin Clark (2010) nomment des « environnements d'apprentissage passifs », c'est-à-dire des dispositifs qui n'acceptent pas d'interactions ouvertes de la part de leurs utilisateurs. Comme le signale Reif (2008) ce type d'environnement peut en partie guider l'apprenant à la place de l'enseignant (tuteur cognitif).

Présentation rapide du cours

Ce cours a trois grandes parties. La première, contient des éléments théoriques pour mieux comprendre le type de tâches d'apprentissage et de feedback (rétroactions) pouvant être réalisées par ordinateur. La deuxième partie, plus pratique, contient les consignes pour réaliser un cours et des QCM avec le logiciel *eXeLearning* ou *Oppia*, le but étant d'appliquer les prescriptions de la première partie. La troisième partie est entièrement dédiée à l'organisation du cours du MARDIF (atelier TICE).

1re partie : Eléments théoriques

La première partie de ce cours présente quelques points de vue théoriques sur le type de tâches d'apprentissage, les principales phases de rétroaction et les formes de feedback pouvant être délivrées de manière (semi-) automatique pour guider l'apprentissage, dans un contexte d'enseignement à distance.

1. Le choix de la tâche

Avant tout, l'une des premières questions que tout enseignant a à se poser lorsqu'il travaille à l'évaluation de ses apprenants porte sur le type de tâche d'apprentissage (ou d'ailleurs, d'évaluation) qu'il va pouvoir leur donner. Il est ici question d'avoir un aperçu général du type de tâches qu'on peut donner, dans un contexte d'enseignement à distance, à des apprenants (bien que cela vaille en grande partie aussi pour des enseignements en présence). Le Tableau 1 ci-dessous détaille le nom des tâches, les processus cognitifs préférentiellement engagés dans leur mise en œuvre, et la conception de l'apprentissage sous-jacente à chacune.

Tableau 1 – Type de tâches, processus cognitifs engagés et conceptions de l'apprentissage sous-jacentes (Trausan-Matu et al., 2008, p. 14).

Tâche	Processus cognitifs	Conception de l'apprentissage
Répondre à un QCM	Apprentissage par cœur. Reconnaissance. Deviner la réponse adéquate.	Transmissive
Rappeler des faits (oral ou écrit)	Apprentissage par cœur	Behavioriste. Implique un rappel simple
Prendre des notes	Traduction d'idées	Transmissive
Ecrire un journal	Traduction d'idées. Construction de connaissances	Constructiviste. Processus méta-cognitifs
Discuter, argumenter à propos d'un problème	Argumentation. Résoudre des problèmes. Trouver des faits (in-)compatibles	Constructiviste
Ecrire un résumé	Traduction d'idées du cours	Transmissive. Implique la compréhension du texte résumé.
Ecrire une dissertation	Traduction d'idées du cours ; mise en œuvre d'un schéma argumentatif spécifique.	Transmissive ou constructiviste (dépend de la tâche donnée)
Réaliser un portfolio	Traduction d'idées du cours, révision, sélection et évaluation de textes, construction de connaissances	Constructiviste. Processus méta-cognitifs

Ce tableau ne couvre bien évidemment pas toutes les tâches d'apprentissage (ou évaluatives) possibles, bien qu'il en soit un utile récapitulatif — par exemple, il omet l'idée de faire concevoir un QCM par les apprenants eux-mêmes, ce qui pourrait être source d'apprentissage, comme le montre Fellenz (2004). Il alerte le lecteur sur un point important : toute activité d'enseignement comprend, même de manière non-intentionnelle, la mention des principes éducatifs qui ont pu présider à sa conception (voir Dessus, 2000, pour une discussion de cet argument).

TP

Choisissez l'une des tâches ci-dessus et décrivez-en une mise en œuvre dans une situation d'enseignement/apprentissage de votre choix. Intéressez-vous particulièrement aux processus cognitifs des apprenants et réfléchissez à la manière d'assister par informatique cette activité (pour qu'elle soit utilisée dans un contexte d'enseignement à distance).

Voir plus loin

L'article de Basque et Lundgren-Kayrol (2002), accessible en *preprint* à http://tecfa.unige.ch/tecfa/teaching/riat140/0304/typologies.pdf est une bonne vue d'ensemble des différentes typologies d'usage des TICE.

2. Les rétroactions par informatique : Quelques phases principales

Les différentes phases d'un feedback pour l'apprentissage

L'informatique, par le biais de nombreux logiciels, procure à l'apprenant de très nombreux moyens d'apprendre et de très nombreux types d'interactions (de Vries, 2001), parfois, sans que leurs utilisateurs s'en rendent vraiment compte. En guise d'illustration, prenons le simple cas des correcteurs orthographiques maintenant intégrés dans tout logiciel de traitement de textes sophistiqué. Utilisons ce cas pour montrer les différentes phases de l'interaction utilisateur-machine.

1. Phase de production : On tape du texte dans une fenêtre du logiciel.
2. Phase de rétroaction : Les mots possiblement mal orthographiés sont mis en évidence, soulignés, par exemple [rétroaction].
3. Phase de correction automatique : On peut, de manière optionnelle (clic droit), se voir proposer des alternatives au mot possiblement mal orthographié. La reconnaissance du mot correcte peut se faire (ou pas).
4. Phase d'action en retour : On peut décider – de laisser le mot dans sa forme originelle (si on est sûr de son orthographe, si le mot est un nom propre, etc.) ; – de sélectionner l'une des alternatives proposées en 3 ; – de corriger directement le mot mal orthographié.
5. Retour en 1, pour taper la suite du texte.

Bereiter (2002), dans le chapitre 2 de son ouvrage (dont un *preprint* est accessible à http://www.cocon.com/observetory/carlbereiter/chapter2.pdf) évoque une procédure simple pour comprendre de quelle manière le correcteur orthographique d'un logiciel de traitement de textes évolué suggère des mots alternatifs à un mot possiblement mal tapé. Lisez le contenu du Tableau 2 ci-dessous et tirez-en quelques déductions sur la manière dont le logiciel en question fonctionne. Vous pouvez vous reporter directement à l'article (en anglais) pour comparer vos conclusions avec celles de Bereiter.

Tableau 2 – Un tableau de suggestions de mots mal orthographiés (Bereiter, 2002, pp. 37–38).

Mots écrits dans le logiciel	Les suggestions du correcteur orthographique
pyschology	psychology, psychologies, psychologic, psychological
pyschilogy, pyschalogy	psychology, psychologic
pyschiology	pas de suggestion
dsychology	pas de suggestion
sychology	psychology, psychologies
cychology	cychology cytology

Cette séquence de « dialogue » entre un utilisateur et la machine est maintenant tellement banale qu'on y participe sans même y penser. Il correspond d'assez près aux modèles du feedback que l'on trouve dans la littérature de recherche. Par exemple, Bangert-Drowns *et al.* (1991, p. 217) ont décrit un modèle cyclique du feedback de six étapes, qui sont les suivantes :

1. Etat initial, c'est-à-dire les caractéristiques cognitives initiales de l'apprenant (motivation, but, connaissances initiales, etc.)
2. Une question est posée à l'apprenant

3. Stratégies de recherche et récupération, c'est-à-dire, les processus cognitifs activés par la question.
4. Réponse à la question
5. Un feedback est délivré à l'apprenant.
6. Evaluation du feedback, l'apprenant peut maintenant évaluer la réponse, en fonction de son degré de confiance et des indices donnés par le feedback. Si l'apprenant est sûr de sa réponse, il est plus à même de mettre en œuvre un traitement engageant des processus cognitifs de haut niveau (résolution de problèmes). En effet, voici ce qui se passe. Si l'apprenant est sûr de sa réponse et que le feedback le confirme, cette réponse sera renforcée. Si, en revanche, il est sûr de sa réponse et que le feedback indique une erreur, il va plus être en position de travailler pour essayer de trouver la source de cet écart.
7. Des ajustements peuvent être faits à l'état initial, qui devient l'état en cours (voir point 1).

TP

1. En utilisant un logiciel de traitement de texte évolué (p. ex., *OpenOffice*), répliquez l'étude de Bereiter (2002) avec quelques mots dont vous ferez varier l'orthographe.
2. Représentez graphiquement les étapes du raisonnement décrit dans l'item 6 ci-dessus (Bangert-Drowns *et al*., 1991).

Voir plus loin

L'article de Dessus et Lemaire (2004) accessible à http://webu2.upmf-grenoble.fr/sciedu/pdessus/dunod06.pdf détaille (dans l'introduction) les différents niveaux d'analyse d'un texte libre et quelles peuvent en être les rétroactions.

3. Les rétroactions : guidage et évaluation

L'étude des rétroactions à donner aux apprenants est un vaste (et ancien) problème, qui n'a pas attendu l'arrivée de l'informatique. Ces rétroactions interviennent à plusieurs niveaux (Enard, 1970) :

- dans le guidage de l'activité de l'élève, lorsque les rétroactions se situent pendant l'activité ;
- dans la connaissance évaluative de sa performance.

Les activités d'évaluation (comme celles de production, voir la section précédente) pouvant être assistées par informatique sont très nombreuses, mais dépendent du type de pédagogie mise en œuvre par l'enseignant (Erstad, 2008). Dans une perspective behavioriste, l'essentiel est que l'élève s'exerce avec de nombreux problèmes répétitifs et les réussisse selon des critères bien définis. Des questionnaires à choix multiple ou bien des tests associant questions et réponses sont dans cette perspective. Dans une perspective constructiviste, l'élève est vu comme plus actif, réalisant des activités de haut niveau (rédaction de textes libres, construction de portfolios), dans des contextes authentiques (p. ex., projets, enquêtes), et réfléchissant à ses performances.

La majorité des logiciels existants s'insèrent beaucoup mieux dans une pédagogie behavioriste, principalement parce que le type de retours (rétroactions) donnés par ordinateur correspondent bien à ses principes (voir Tableau 1), bien qu'il existe de plus en plus de logiciels permettant un travail dans une perspective constructiviste (mais pas toujours évalués par le logiciel lui-même).

Hattie et Timperley (2007) montrent que les retours peuvent porter sur les trois types d'éléments. Là encore, il est possible d'utiliser cela pour guider les élèves.

- où vais-je ? (quels sont les buts que je poursuis ?) ;
- comment y vais-je ? (quels progrès ai-je réalisés vers ce but ?) ;
- quoi faire ensuite ? (quel travail dois-je faire ensuite ?).

TP

Prenez un logiciel éducatif de votre choix et classez les différents types de feedback qu'il propose en utilisant les 3 questions de Hattie et Timperley (2007). Ces questions peuvent également être utilisées pour guider la circulation d'un apprenant au sein d'un logiciel d'apprentissage ou une plate-forme d'enseignement à distance.

Pour en savoir plus

Le chapitre 2 (pp. 6–20) de Luengo (2009) disponible à http://membres-liglab.imag.fr/luengo/textes/articles/ToutHDRV2Luengo.pdf

4. Différentes formes de feedback

On catégorise usuellement les formes de feedback selon leur intention pédagogique, leur complexité, l'agent délivrant le feedback (la suite de cette section est issue et traduite de Trausan-Matu et al., 2008, p. 20 et sq.).

Feedback directif vs. Facilitatif

Shute (2008) a proposé la distinction entre un feedback directif et facilitatif. Un feedback directif a pour but d'informer l'apprenant sur ce qu'il a à faire, alors que le feedback facilitatif guide l'apprenant dans sa compréhension du contenu. Cette approche facilitative est un type de feedback « étayage », dont les caractéristiques ont été définies par Bransford *et al*. (2000) :

- augmente l'intérêt de l'apprenant pour la tâche ;
- simplifie la tâche pour qu'elle soit plus aisément mise en œuvre ;
- donne des pistes afin d'aider l'apprenant à accomplir le but recherché ;
- indique clairement l'écart entre le travail de l'élève et la solution standard ;
- réduit la frustration et le risque ;
- définit clairement ce qui est attendu dans l'activité d'apprentissage à réaliser.

L'avantage principal d'un feedback facilitatif est qu'il procure un guidage et des indications pour améliorer l'apprentissage. Pour un apprenant novice, un feedback directif est plus utile, alors qu'un feedback facilitatif est plus judicieux pour les apprenants plus avancés. Cela étant, le rôle facilitateur du feedback peut ne pas jouer si (Kulhavy, 1977) :

- il est disponible aux apprenant *avant* qu'ils puissent répondre, ce qui les amène à copier la réponse plutôt qu'à lire et comprendre le contenu ;
- le feedback et/ou le contenu est trop complexe pour les apprenants, ce qui les amène à passer du temps à essayer de les comprendre.

La complexité du feedback

Shute (2008) a réalisé une revue complète des formes de feedback classés par degré de complexité croissante. Le Tableau 3 ci-dessous les résume.

Les points ci-dessous nous montrent qu'il y a la place pour la conception d'exercices, notamment sur support informatisé, procurant un retour immédiat à l'élève. Ces exercices peuvent être appelés « exercices d'application » ou « d'entraînement ». Lemercier *et al.* (2001) distinguent les exercices d'application, tâche que l'on propose après un cours où l'on a exposé une notion (savoir) ou une procédure (savoir-faire) où il s'agit d'appliquer ce que l'on a compris ; des exercices d'entraînement, où les élèves travaillent à automatiser des procédures en cours d'acquisition (à les automatiser). Enfin, il existe également des exercices résolus, montrant pas à pas une procédure (p. ex., de résolution de problème, un savoir-faire) jusqu'à sa solution, et où des feedback sont fournis à chaque étape.

Tableau 3 – Formes de feedback classés par complexité croissante (Shute, 2008, p. 160).

Forme de feedback	**Description**
Aucun feedback	Aucune indication sur l'exactitude de la réponse de l'apprenant n'est donnée.
Vérification	Egalement appelé « connaissance des résultats ». Informe l'apprenant sur l'exactitude de sa réponse (p. ex., vrai-faux, ou un pourcentage global).
Réponse correcte	Egalement appelé "connaissance de la réponse correcte". Informe l'apprenant de la réponse correcte à un problème spécifique, sans information additionnelle.
Essayer encore	Egalement appelé le feedback « répéter tant que faux ». Informe l'apprenant d'une réponse incorrecte et lui permet un ou plusieurs essai(s) supplémentaire(s).
Mise en évidence des erreurs	Egalement appelé « surlignage des erreurs ». Les erreurs sont mises en évidence, sans donner la réponse correcte.
Elaboré	Terme générique qui signale qu'on a donné une explication sur la raison pour laquelle une réponse donnée était correcte ou non. Pourrait permettre à l'apprenant de revoir une partie de ce qui a été enseigné. Peut ou pas présenter la réponse correcte (voir plus bas les 6 autres types de feedback élaboré).
Elaboré avec attributs critiques	Feedback élaboré qui présente à l'apprenant les caractéristiques les plus importantes du concept ou de l'habileté) étudié, pour une meilleure compréhension (attributs critiques).
Elaboré relatif au contenu étudié	Feedback élaboré qui présente à l'apprenant des informations relatives au contenu étudié (p. ex., de la réponse correcte). Peut impliquer la répétition de l'enseignement.
Elaboré relatif à la réponse	Feedback élaboré qui se focalise sur la réponse de l'apprenant. Peut décrire pourquoi la réponse incorrecte est fausse, ou la réponse correcte est juste. Ce feedback ne recourt pas à une analyse des erreurs.
Elaboré sous forme d'indications ou signaux	Feedback élaboré guidant l'apprenant dans la bonne direction (p. ex., indications sur la stratégie à utiliser ensuite, exemple résolu, démonstration). Evite explicitement de présenter la réponse correcte.
Elaboré, centré sur les erreurs ou compréhensions erronées	Feedback élaboré qui nécessite une analyse et un diagnostic des erreurs possibles. Il donne à l'apprenant des informations sur ses erreurs ou compréhensions erronées (p. ex., ce qui est faux, et pourquoi).
Informatif	Le feedback le plus élaboré (d'après Narciss & Huth, 2004), est un composite de feedback de vérification, indications et indications stratégiques sur la manière de procéder.

TP

En reprenant le contenu du Tableau 3 ci-dessus, et en prenant un contenu de cours et une question à propos de ce dernier, essayez formuler un feedback de ce cours selon 3 ou 4 différentes formes.

Voir plus loin

Lire le rapport de Lemercier *et al.* (2001), accessible à http://edutice.archives-ouvertes.fr/docs/00/00/16/38/PDF/EXTricot.pdf

2e partie : Les questionnaires à choix multiple

La deuxième partie de ce cours présente quelques éléments de la recherche en psychologie de l'éducation sur les questionnaires à choix multiple et se termine par quelques éléments pratiques permettant de réaliser de tels questionnaires qui évitent les biais les plus courants, et détaille la méthode de travail de l'activité 2.

5. Les questionnaires à choix multiple : bref historique

La première utilisation en vraie grandeur d'un QCM a eu lieu il y a exactement 100 ans. Madaus et O'Dwyer (1999) datent l'utilisation des premiers questionnaires à choix multiple (QCM) au tout début du XXe siècle, pour pallier à la fois la mauvaise validité des essais, couramment utilisés comme moyen d'évaluation et dont l'évaluation était jugée trop subjective (voir Davidson, 2011; Leclercq, 2006), mais aussi pour rendre la correction plus efficiente : cela prend beaucoup de temps de corriger un essai.

Les premiers tests QCM apparaissent en 1915, sous l'impulsion de Frederick J. Kelly, de l'Institut de formation des enseignants du Kansas (Davidson, 2011), qui crée le premier test de lecture silencieuse chronométré. L'essor est rapide : en 1921, 2 millions de soldats américains et 3 millions d'élèves passent un QCM (Monaham, 1998). Un peu plus tard, en 1926, une version de ce test est introduit dans le SAT (*Scholastic Aptitude Test*, évaluation des connaissances des élèves, de l'école primaire au lycée, aux États-Unis d'Amérique), et ont fleuri dans le cadre de tests standardisés, quand ont crû la demande, et surtout le nombre d'étudiants et d'élèves.

En 1955, l'invention du scanneur optique a permis de corriger très rapidement ces formulaires, ce qui a accentué la popularité de ces outils, ensuite augmentée encore par l'arrivée de l'ordinateur, dans le courant des années 1970. Cela rendait l'évaluation très rapide et peu chère (résultats en 2 à 4 semaines, pour un coût de 2 à 3 dollars lorsque réalisée via scanneur). Les années 1980, comme l'expliquent Madaus et O'Dwyer, ont vu arriver un regain de critiques envers les QCM, jugés (avec raison) comme insuffisamment authentiques, et/ou donnant des résultats non fiables. Plus récemment encore, dans le cadre de l'e-learning, puis de l'utilisation de boîtiers de vote en classe, les QCM sont redevenus plus utilisés. Notre opinion est qu'un test correctement construit et évalué peut donner des résultats aussi fiables que d'autres méthodes.

Cet intérêt et besoin de tests a lancé une industrie très opulente, et il ne rentre pas dans le propos de ce cours de la relater plus avant (voir Clarke, Madaus, Horn, & Ramos, 2000).

Voir plus loin

Lire l'article de Leclercq (2006), disponible à http://www.elearning.ulg.ac.be/evolution_qcm_in_figari_mottier.pdf

Lire l'article de la Wikipédia sur le choix multiple : http://en.wikipedia.org/wiki/Multiple_choice

6. Les questionnaires à choix multiple : définitions et formes

La technologie de l'écriture des QCM a pris son essor dans les années 1970, quand il s'est agi de proposer des tests d'apprentissage qui mesurent plus fidèlement l'évolution de ce dernier, voir Roid et Haladyna (1980) pour une revue. Nous avons choisi de focaliser le reste de ce cours sur la conception et réalisation de questionnaires à choix multiple (que nous distinguerons de questions vrai/faux). La raison première est qu'ils sont maintenant très répandus dans le domaine de l'enseignement à distance (et parfois même en présence).

Commençons par quelques définitions. Un questionnaire à choix multiple (ou test) est composé de plusieurs questions, ou items de test. Osterlind (1998, p. 18) définit un item de test de la manière suivante : « […] une unité de mesure avec un stimulus et une forme de réponse prédéfinie, qui est conçue pour solliciter une réponse d'une personne testée, à partir de laquelle un construit psychologique donné (connaissance, habileté, prédisposition, trait) peut être inféré. » La réponse fait l'objet d'un score intermédiaire (0 en cas de mauvaise réponse, 1 en cas de bonne réponse) et les valeurs intermédiaires sont agrégées pour obtenir un score global. De plus, avant les questions et réponses, un bref descriptif du contexte peut être ajouté.

Plus précisément, un item de question à choix multiple comporte : – une amorce (question) ; – la réponse correcte ; – des réponses distractrices. Haladyna (2004) signale que la question doit donner une idée, en elle-même, du type de connaissance à sélectionner. La réponse correcte (pas nécessairement une phrase, mais aussi un symbole, un diagramme, etc.) doit l'être indiscutablement. Les distracteurs, par contraste, doivent être indiscutablement des réponses erronées, mais à la fois plausibles et difficilement devinables, ce qui est plus coûteux à déterminer pour l'enseignant concepteur.

On a souvent critiqué, et à juste raison, les QCM pour le problème du choix au hasard (ce qui suit est issu de Bar-Hillel *et al*. (2005) : un élève qui répond à un item a une probabilité non nulle de répondre juste à la question, même s'il ne connaît pas la bonne réponse. De ce fait, il est impossible pour un enseignant de distinguer les réponses justes « par connaissance » des réponses juste « par chance » (et aucune formule de calcul ne permet de le faire : on peut juste essayer de *décourager* les choix au hasard). Il arrive justement de pénaliser les réponses erronées pour décourager le choix au hasard (mais, *a contrario*, il est impossible de distinguer une omission liée à l'ignorance plutôt qu'à l'aversion au risque). La question de la « devinabilité » des items est donc importante, mais difficile à résoudre (voir Hanna & Oaster, 1978, pour une synthèse).

Toutefois, un élève cochant au hasard, d'une part, n'est pas si fréquent et, d'autre part, n'aurait que peu de chances d'avoir une note acceptable, si l'on se fie à un simple calcul de probabilités. Un élève a en réalité, plutôt, une connaissance *partielle* du domaine, ce qui l'amène à écarter certaines réponses moins plausibles et lui donne une probabilité d'avoir une bonne note non négligeable (Burton, 2005). En d'autres termes, il faudrait récompenser les élèves qui utilisent une telle connaissance partielle plutôt que ceux devinant les distracteurs grâce à un test mal conçu.

Il convient donc, pour éviter cela, soit de concevoir un QCM où aucune réponse fausse n'est spécialement devinable comparée aux autres (par exemple, en plaçant la bonne réponse aléatoirement parmi les distracteurs), soit d'amener l'élève à un calcul ou un travail intermédiaire.

En effet, un QCM qui n'amène pas de traitement particulier à l'élève aura une réponse plus devinable avec du bon sens (ou le simple lien entre la question et la réponse) qu'un QCM qui

amène à réaliser un traitement particulier (p. ex., calcul, réflexion). Dans ce dernier cas, les élèves ne pourront aisément deviner quelle est la bonne réponse, comme ci-dessous, où seul le calcul amène à R3 :

Q : Quel est la circonférence d'un cercle de diamètre 3 cm.

R1 : 9,1 cm

R2 : 9,3 cm

R3 : 9,4 cm

Cela est bien sûr plus difficile, bien que non impossible, pour des connaissances de type « sciences humaines » ou de culture générale. Par exemple, dans la question suivante (issue de http://www.makem.co.uk/index.php/surgery-hour/79-how-to-write-qorking-quiz-questions?showall=&start=5), toutes les villes ont des graphèmes plausibles, et aucune n'est trop connue (Bucarest n'est pas mentionnée) :

Q : Parmi les villes suivantes, quelle sont celles qui sont en Roumanie ? 1) Odessa; 2) Oradea; 3) Arad; 4) Nesebâr; and 5) Timişoara

R1 : 2, 3 et 5

R2 : 1, 2, et 4

R3 : 1, 3, 5

Pour répondre à certaines critiques faites aux QCM (notamment, le fait qu'on pioche une réponse parmi d'autres, sans réelle compréhension), il est tout à fait possible que l'amorce ne soit pas qu'une question amenant une sélection (*e.g.*, « Sélectionnez la bonne réponse »), mais peut amener le questionné à réaliser une tâche plus complexe (voir section suivante pour d'autres précisions).

D'autres formes plus évoluées de QCM ont été proposées, dont les ARQ (*assertion-reason questions*, questions assertion-raison). Ce type de questionnaire est censé impliquer l'apprenant dans des activités de plus haut niveau. Williams (2006) en décrit la construction de la manière suivante (voir aussi le site http://www.caacentre.ac.uk/resources/objective_tests/assertion.shtml). La question est en deux parties : une assertion et une raison, et l'apprenant doit déterminer, dans un premier temps si chacune des deux est vraie ou fausse. Dans le cas où les deux sont vraies (et seulement dans ce cas-là), l'apprenant détermine ensuite (en répondant à une deuxième question), si la raison donnée est bien une explication correcte de l'assertion. Pour chaque couple d'assertion-raison, cela fait donc 5 réponses possibles (A est l'assertion, R la raison) :

1. A et R sont vraies et R est l'explication correcte de A ;
2. A et R sont vraies et R n'est pas l'explication correcte de A ;
3. A est vraie et R est fausse ;
4. A est fausse et R est vraie ;
5. A et R sont fausses.

Voici un exemple d'assertion-raison et la réponse :

Tableau 4 – Un exemple de question Assertion-raison et sa solution.

Assertion	Raison
Les responsables de la centrale de Tchernobyl ont réalisé un exercice pour prouver qu'on pouvait relancer la centrale avec des générateurs diesel à la suite d'une panne totale du réseau électrique, ce qui a déclenché la catastrophe.	Le réacteur s'est trouvé en sous-puissance un bref temps et dans une grande instabilité, ce qui a provoqué une explosion. Il aurait dû être refroidi plusieurs jours pour que l'iode et le xénon se désintègrent.
Réponse : A : V ; R : V ; R est l'explication de A (soit la réponse 1) source : http://fr.wikipedia.org/wiki/Catastrophe_de_Tchernobyl (accédée le 7 avril 2011).	

TP

Construisez sur papier un item de test de type assertion-raison.

Voir plus loin

Il existe deux bonnes synthèses accessibles sur internet sur les possibilités d'informatiser un QCM à des fins éducatives : Bravard (2005), et Seyve et Grépilloux (2004), respectivement à http://fle.u-strasbg.fr/evaluation_fle/Bravard_qcm.pdf et à http://greco.grenet.fr/documents/dossier_greco_qcm.pdf

7. Les aspects cognitifs des QCM

Les tests via QCM, s'ils sont maintenant très répandus à tous niveaux scolaires, souffrent de deux préjugés que nous allons détailler ici. Le premier est que les QCM ne peuvent mesurer que des compétences de bas niveau dans l'échelle de Bloom *et al.* (1969) (p. ex., rappeler des faits). Le second, que le fait de répondre à un test ne fait que *prélever* de l'information sur ce que peut connaître un apprenant sans aucunement interagir avec cette connaissance ni l'affecter.

Les effets bénéfiques des QCM sont les suivants (Marsh, Roediger, Bjork, & Bjork, 2007) : ils donnent l'occasion d'exercer sa mémoire (reconnaître des informations), d'étudier un contenu, et les différents indices donnés dans les réponses de choix sont traités en tant que problèmes, ce qui peut, là encore, améliorer l'apprentissage, de plus, comme ils ont rapidement connaissance de ce qu'ils ne maîtrisent pas, ils peuvent en entreprendre l'étude. Du point de vue de l'enseignant, ils sont aisés à corriger, donc informatisables souvent utilisés quand les cours ont un grand nombre d'étudiants. Parmi les inconvénients connus des QCM : ils sont difficiles à réaliser, tout apprenant peut, en répondant au hasard, espérer avoir une note acceptable (à tout le moins, non nulle), ils induisent également l'apprentissage de mauvaises réponses, puisqu'elles sont traitées cognitivement par l'apprenant. Et, bien sûr, il n'existe pas de QCM dans le monde réel, ce qui rend son usage assez peu authentique (Williams, 2006). Du point de vue de l'enseignant, également, ils sont assez difficiles à construire, puisque les réponses distractrices doivent être à la fois plausibles et faux (voir Section 6).

Les processus cognitifs engagés dans les QCM

Il est habituellement considéré que la réponse aux QCM n'incite pas les apprenants qui les formulent dans des processus cognitifs de haut niveau, puisque nécessairement fondée sur la *reconnaissance* d'informations plutôt que leur *formulation* (puisque les réponses sont nécessairement formulées par avance pour être choisies) ; le fait qu'on puisse y répondre totalement au hasard est en grande partie responsable de cela (voir Leclercq, 2006 pour une discussion de ce point). Toutefois, comme l'ont signalé Roid et Haladyna (1980), il est possible de centrer des QCM sur d'autres activités que la simple sélection de faits. Ils distinguent, en adaptant la taxonomie des objectifs cognitifs de Bloom et ses collègues, 5 niveaux (voir Tableau 5, qui montre le type d'amorces amenant des activités plus complexes qu'une simple sélection-reconnaissance de réponse).

De plus, Marsh *et al*. (2007) reportent que les effets d'un QCM sur l'apprentissage sont aussi importants lorsque le QCM porte sur des aspects définitionnels (niveau 1 de la taxonomie de Bloom) que sur des aspects d'application (niveau 5 de la définition). Un travail plus récent (Tractenberg, Gushta, Mulroney, & Weissinger, 2013) montre même qu'il peut exercer les niveaux les plus élevés de la pensée critique.

Les QCM font-ils apprendre ?

Un avantage peu mis en avant, et pourtant évoqué dans la littérature de recherche est qu'il améliore l'apprentissage : des apprenants confrontés à un apprentissage et passant un test à son propos ont de meilleures performances que des apprenants qui réalisent l'apprentissage sans passer le test (Bangert-Drowns et al., 1991). Cela est dû à un effet bien connu dans le domaine de l'étude de l'apprentissage, l'effet de test (*testing effect*) : une information rappelée de la mémoire à long terme, par exemple lors d'exercices d'entraînement, sera plus aisément

retrouvable (activable) par la suite, par exemple lors de l'examen final (McDaniel, Anderson, Derbish, & Morrisette, 2007).

De nombreuses recherches ont eu pour but de valider, d'un point de vue cognitif, la passation de QCM auprès d'apprenants. Roediger et Butler (2011) en font une synthèse intéressante. Le résultat le plus frappant est de montrer que, contrairement à ce qui est communément pensé, le simple fait d'exposer des apprenants à un test leur fait apprendre un peu plus le matériel. Ces auteurs ont réalisé une expérimentation (Butler & Roediger, 2008) dans laquelle des apprenants devaient étudier un matériel à apprendre (un cours d'histoire sur les Khmers rouges), puis étaient confrontés à un test de compréhension final. Puis ils étaient répartis en 4 groupes : sans test (aucun test intermédiaire) ; test intermédiaire sans feedback ; test intermédiaire avec feedback (donnant les réponses correctes) ; test avec feedback reporté à la fin du test. Les résultats sont décrits dans la Figure 1 ci-dessous.

Tableau 5 – Un exemple de typologie cognitive de l'écriture d'items de QCM (Roid & Haladyna, 1980, p. 299).

Activité cognitive	Définition	Formes	Exemple de question-amorce
Rappeler des faits	La reproduction d'un élément-stimulus identique à celui présenté	Nommer, Sélectionner, Décrire,	À quelle date Christophe Colomb a-t-il découvert l'Amérique ? a. 1492 ; b. 1489 ; c. 1776
Résumer	La compréhension des concepts et la capacité à identifier correctement les exemples, instances ou attributs du concept	Identifier, Définir, Traduire	Quel est un bon exemple d'allitération [en anglais] ? a. *gurgling* ; b. *school–pool* ; c. *blue–blood* ; d. *up–down*
Prédire	Utiliser des règles pour mettre au jour des relations de contingence entre items d'apprentissage. L'apprenant lit une situation et peut anticiper une conséquence fondée sur une règle.	Si… alors…	Si la température du fluide dans la bouteille excède 100°C, alors, a. tous les fluides s'évaporeront b. le liquide explose c. rien ne se passe
Evaluer	La capacité à a) sélectionner un ou des critère(s), b) utiliser un critère, c) sélectionner et utiliser un critère pour mettre en œuvre une décision.	Quel item est le meilleur, le pire, le plus, le moins, etc.	Du point de vue de l'efficacité, quelle procédure est la meilleure ? a. procédure 1 b. procédure 2, etc.
Appliquer	Résolution d'un problème impliquant a) le comprendre ; b) le définir ; c) choisir des principes, règles ou méthodes qui permettront de le résoudre ; d) choisir ou trouver des solutions.	Pas de formes standard.	Quel plan permettra de réaliser un système de distribution urbaine de l'eau qui sera efficace en toutes saisons ? a. nappe phréatique à l'ouest de la ville b. nappe phréatique à l'est de la ville c. un réservoir dans les collines d. un pipeline

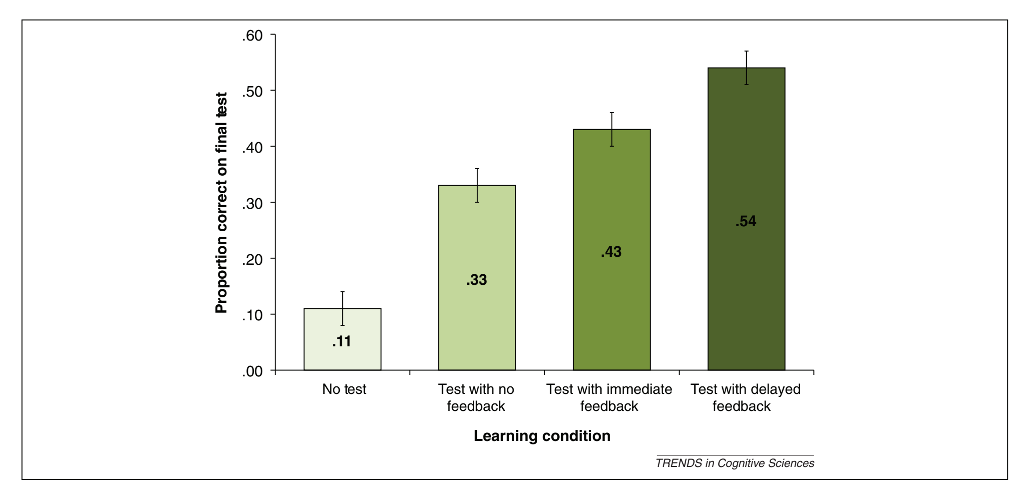


Figure 1 — Les 4 conditions d'apprentissage de l'expérience de Butler et Roediger (2008), copié de Roediger et Butler, 2011, p. 23.

Les QCM sollicitent principalement une activité de reconnaissance (reconnaître la bonne réponse parmi plusieurs). Cela a un inconvénient majeur : les apprenants doivent lire des réponses erronées, et le danger est grand qu'ils les apprennent (ce que les études de Butler ont confirmé). Pour pallier ce risque, il convient de donner, le plus tôt possible après la passation

du QCM, les bonnes réponses à l'apprenant. Sadler (1999) a également montré qu'il est possible de construire des QCM en indiquant, comme choix possibles, les compréhensions erronées des apprenants.

Montepare (2005) a mis au point une méthode d'auto-correction des QCM intéressante. Les QCM sont passés en utilisant des stylos d'une couleur donnée et, au lieu d'être corrigés tout de suite, les élèves s'engagent (seul ou collaborativement, le cas échéant avec le recours aux notes de cours et/ou manuels) dans une phase d'auto-correction où ils reprennent chaque item et déterminent si leur réponse était bien la bonne. S'ils estiment que non, ils signalent leur nouveau choix d'une autre couleur. Ensuite, l'enseignant corrige les QCM en attribuant le score de 1 si le choix initial était correct, de 0,5 si c'est le second choix qui était le bon, et de 0 dans les autres cas. L'auteur a remarqué un meilleur engagement de ses élèves avec cette méthode. Une expérimentation contrôlée des effets de cette méthode a récemment été réalisée et donne des résultats mitigés (Francis & Barnett, 2012).

Une autre étude (Bleske-Rechek, Zeug, & Webb, 2007) a comparé les performances d'étudiants en psychologie, selon qu'ils répondaient à un QCM ou composaient de courtes réponses ouvertes à des questions d'examen. Ils ont montré qu'il y a une corrélation entre leurs résultats QCM vs. réponses courtes, ensuite, ils montrent que les scores des étudiants au QCM sont mieux reliés à leurs scores de réussite universitaire générale que ceux aux réponses courtes. Il semblerait donc que les performances des étudiants à des QCM aient un lien plus profond avec leur performances académiques que celui souvent estimé par ses détracteurs.

Enfin, une étude (McDaniel, Wildman, & Anderson, 2012) a même montré que faire passer un QCM avec correction immédiate à des étudiants à des fins de révision avant un contrôle donnait les mêmes résultats que leur faire répondre à des questions ouvertes et courtes, et des résultats supérieurs à la simple relecture du cours. Point plus important, les étudiants réussissent également mieux à répondre à des questions *reliées au cours*, mais non similaires (p. ex. une question sur l'information arrivant dans une structure du cerveau *vs*. en sortant). Cela peut s'expliquer de deux manières (Bjork, Little, & Storm, 2014) : soit les étudiants révisent efficacement les lacunes pointées par les rétroactions du QCM, soit une réflexion à propos des différentes réponses possibles les amène à approfondir les relations entre ces réponses, et donc à apprendre.

Toutefois, un rapport (Means, Toyama, Murphy, & Bakia, 2010) montre des résultats plus mitigés à propos de l'insertion de QCM dans des cours à distance : elle n'est pas nécessairement en lien avec un meilleur apprentissage des étudiants.

TP

Trouver d'autres avantages et d'autres inconvénients des QCM pour évaluer l'apprentissage. Aller sur la page wikipedia à propos des questionnaires à choix multiple : http://fr.wikipedia.org/wiki/Questionnaire_à_choix_multiples. La recopier dans un fichier traitement de textes et l'améliorer en citant ses sources.

Voir plus loin

Lire l'un des manuels de construction de QCM cités en références, pour plus d'informations.

8. Les types de QCM : choix et conseils de rédaction

Cette section a pour but de donner des conseils plus précis sur la manière de concevoir des QCM, et de choisir leur forme en fonction des buts d'enseignement. Haladyna *et al.* (2002, p. 322 et sq.) listent de manière exhaustive les types de QCM qu'il est possible de réaliser.

Choix multiple conventionnel

Choisir la réponse qui décrit le plus clairement le processus de polinisation.

A. La rencontre d'une cellule œuf et d'une cellule spermatozoïde.

B. Le transfert de graines de pollen dans le pistil.

C. De la nourriture est consommée et de l'énergie est délivrée.

Choix alternatif

Quel est la réponse parmi les deux suivantes qui ralentit le plus efficacement le processus de respiration des plantes.

A. Temps froid

B. Temps orageux

Vrai-Faux

La capitale de l'Uruguay est Montevideo.

Vrai-Faux multiple

Vous êtes un fermier expert en matière organique. Vous connaissez les secrets pour faire pousser les plantes. Quels sont les réponses, parmi les suivantes, qui décrivent vos tours de main ? (cocher A si vrai, B si faux).

1. Lorsque vous plantez des haricots, vous le faites dans un endroit bien ombragé, afin qu'ils reçoivent peu ou pas de lumière.
2. Lorsque vous semez vos graines, vous les arrosez et vous assurez que le sol reste humide.
3. Vous plantez vos graines uniquement quand la température est appropriée.
4. Pour favoriser la pollinisation, vous répandez de l'insecticide sur vos cultures pour éloigner les abeilles et autres insectes.

Appariement

Apparier chaque terme de droite avec sa description de la colonne de gauche.

1. Attire les abeilles	A. Graines de pollen
2. Produit les graines de pollen	B. Pétales
3. Abrite les cellules œuf	C. Fleur
4. Les graines formées	D. Etamine
5. Contient l'ovaire	E. Ovaire
	F. Pistil

Choix multiple complexe

Parmi les réponses suivantes, lesquels sont des fruits ? Cochez la bonne réponse (lettre).

1. Tomate ; 2. Tomatille ; 3. Piment habanero

A. 1 & 2

B. 2 & 3

C. 1 & 3

D. 1, 2, & 3

Items dépendants du contexte

Imaginez que vous êtes un délégué de l'Etat du Massachusetts à la convention. Vous avez été autorisé à voter au nom de votre Etat.

1. Vous approuveriez sans doute

A. Le plan du New Jersey.

B. Le plan de la Virginie.

2. En lien avec votre expérience acquise dans le traitement des dossiers, vous pensez :

A. que les fermiers n'ont pas à répercuter la charge des taxes aux gens de la ville.

B. que les indiens (américains natifs) ont à être pris en compte dans l'imposition.

C. etc.

Ces mêmes auteurs (p. 312 du même article) proposent une liste de principes guidant la construction de QCM, réalisée en compilant de nombreux articles ou ouvrages sur le sujet. Elle est reportée ci-dessous (traduction libre, et modification de l'ordre des items). Pour rappel : un questionnaire est composé de différents items, eux-mêmes composés d'une amorce et de plusieurs choix de réponse.

Le contenu

1. Chaque item doit couvrir un contenu spécifique et un comportement mental unique, comme dans l'organisation de tests.

2. Faire en sorte que chaque item ait un contenu indépendant de celui des autres (et donc que sa compréhension ne nécessite pas la lecture d'un autre, comme avec des références de type : « aucun des autres », « tous les autres »). En effet, cela transforme la tâche en une évaluation indépendante des réponses en vrai/faux (Towns).

3. Chaque item doit être fondé sur un contenu d'apprentissage significatif ; éviter des contenus triviaux.

4. Eviter les contenus d'items trop spécifiques, ou au contraire trop généraux.

5. Utiliser un matériel nouveau pour tester des apprentissages de haut niveau. Modifier en les paraphrasant les formulations des manuels ou celles utilisées par les enseignants en cours pour éviter les simples rappels de phrases par cœur.

6. Eviter les items exposant des opinions.

7. Eviter les items avec des astuces (comme, par exemple, les négations).

8. Ecrire les items avec un vocabulaire au niveau du public testé.

9. Evitez que l'on puisse deviner la réponse à la seule lecture de la question (c'est-à-dire, sans les connaissances nécessaires, par exemple, lorsque présentées dans un texte à lire juste avant le QCM). Hanna et Hoaster (1978) signalent que les réponses à certains items sont trop aisément devinables (ou, *a contrario*, que d'autres sont trop aisément détectables en tant que distracteurs). Par exemple, si l'élève lit le texte et a à répondre à l'item suivants (*id.*, p. 229) :

T « Les oiseaux volaient au-dessus de l'océan ».

Q. Qui volait au-dessus de l'océan ?

R1 : Des avions

R2 : Des chauve-souris

R3 : Des oiseaux

R4 : Des frelons

Les élèves y répondant ont plus de chances que le hasard d'y répondre correctement, puisque les réponses 2 et 4 ne sont pas suffisamment plausibles *indépendamment de la compréhension de l'élève*.

Le format

9. Ne pas utiliser le format « choix multiple complexe ».

10. Formater les items verticalement et non pas horizontalement.

Le style

11. Relire et tester les items avant passation.

12. Ecrire des phrases correctes du point de vue grammatical, orthographique, mais aussi de la ponctuation et typographie (capitales, etc.)

13. Faire en sorte de réduire le temps de lecture de chaque item.

L'amorce

14. S'assurer que les pistes évoquées dans l'amorce sont claires.

15. Inclure l'idée centrale dans l'amorce plutôt que les choix.

16. Eviter les amorces verbeuses.

17. Eviter les phrases comportant des négations (ne... pas..., excepté) dans les amorces. Si cela ne peut être évité, le faire avec précaution et mettre les négations en évidence (gras, capitales).

Les choix

18. Ecrire autant de choix pertinents que possible, mais des recherches suggèrent que trois choix sont suffisants (Towns).

19. S'assurer que seulement un choix correspond à la réponse correcte.

20. Varier la place de la bonne réponse en fonction du nombre de choix.

21. Placer les choix dans un ordre logique ou numérique.

22. Rendre les choix indépendants les uns des autres : ils ne doivent pas y avoir des références de l'un à l'autre.

23. Rendre la formulation des choix homogène en contenu et en structure grammaticale.

24. Rendre la longueur des phrases des choix à peu près égale.

25. La formule « aucun des choix ci-dessus » doit être utilisée avec précaution.

26. Eviter la formule « tous les choix ci-dessus ».

27. Eviter la formule « tous les choix ci-dessus ».

28. Formuler les phrases positivement. Evitez les négations telles que *pas.*

29. Eviter de donner des indices menant à la bonne réponse, tels que les suivants :

a. Ne pas utiliser des adverbes tels que « toujours », « jamais », « absolument » (ils signalent des réponses très spécifiques, donc peu plausibles).

b. Ne pas faire des associations par la prononciation ou par des mots se ressemblant.

c. Ne pas faire d'incohérences grammaticales qui orientent vers la réponse correcte.

d. Ne pas écrire de réponses correctes trop voyantes.

e. Ne pas écrire des paires ou triplets d'options qui orientent vers la réponse correcte.

f. Ne pas écrire de réponses visiblement absurdes ou ridicules.

30. Rendre plausibles toutes les réponses distractrices.

31. Utiliser les erreurs typiques des apprenants pour écrire les réponses distractrices.

32. Utiliser l'humour seulement s'il est compatible avec les pratiques de l'enseignant et l'environnement d'apprentissage.

33. Si des éléments de l'amorce doivent être repris dans les choix de réponses, faire en sorte de les répliquer dans tous les choix (car leur absence peut être le signal d'une réponse fausse).

9. Construire un cours avec un outil de génération de cours interactifs

La procédure ci-dessous vise à concevoir rapidement un cours diffusable avec un outil de génération de cours interactifs. Les années précédentes de ce cours, seul le logiciel *eXe Learning* était proposé. Cette année, avec l'arrivée d'*Oppia*, système de conceptions de parcours interactifs très puissant, nous avons décidé de laisser aux étudiants le choix de l'outil.

Ces deux logiciels ont la particularité (outre celle d'être libres et gratuits) de permettre de mêler, dans la même page, du contenu multimédia de présentation de contenu *et* d'évaluation de l'apprentissage. Bien évidemment, il est possible d'utiliser la plupart des prescriptions de cette section pour réaliser des QCM sur d'autres logiciels de fonctionnalités similaires (p. ex., *NetQuiz Pro*).

Nous conseillons vivement aux étudiants de commencer à utiliser *eXe Learning*, qui est bien plus simple d'emploi qu'*Oppia*. Les étudiants les plus avancés (et curieux), s'ils buttent sur les limites de ce logiciel, pourront utilement, et avec assez peu de travail supplémentaire, transférer leur contenu sur *Oppia*, et profiter ainsi de fonctionnalités plus élaborées (calcul de scores, prise en compte du parcours de l'élève, traces). Ce cours se déroulant sur 2 ans, la découverte d'*Oppia* peut tout à fait se faire durant la deuxième année, également.

Il n'est pas dans l'objet de ce cours (puisqu'il s'agit d'un cours de Master Recherche) de proposer un mode d'emploi étendu de chacun de ces logiciels. Nous laisserons donc les étudiants se former eux-mêmes. Toutefois, comme il existe très peu de documents sur *Oppia* (et aucun en français), la deuxième partie de cette section comprend une brève description de quelques fonctionnalités de base d'*Oppia*. Il sera nécessaire aux utilisateurs avancés de lire la documentation en anglais pour tirer tout le parti de ce logiciel.

eXe Learning

Il est donc conseillé de suivre pas à pas la procédure ci-dessous, afin de prendre progressivement connaissance des possibilités de travail d'*eXe Learning*. Comme déjà dit, ne passez à *Oppia* uniquement si les deux conditions suivantes sont remplies : 1. vous êtes une personne compétente en informatique ; 2. vous trouvez *eXe Learning* ne possède pas les fonctionnalités nécessaire pour votre cours.

Cette procédure comprend deux parties principales : A. Information sur le logiciel et son installation ; B. Conception du cours et des questions. Un troisième ensemble de manipulations concerne le téléversement des fichiers produits sur un site web. Cet ensemble est décrit dans la troisième partie du document et concerne les étudiants du cours.

A. Prise d'informations sur les fonctionnalités du logiciel, installation et précautions

1. Visionnement du tutoriel vidéo présentant les principales fonctionnalités du logiciel, disponible à http://www.clicattitude.info/tutoexe/dcouvrir_la_fentre_de_travail_dexe_learning.html
2. Visionnement de quelques travaux d'étudiants, afin de comprendre la logique du logiciel ; de ce qu'il permet de faire ou de ne pas faire :
    a. http://perso.numericable.com/annick.pradeau/ViolenceConjugale/index.html (cours conçu par une étudiante de la promotion 2010–11 du MARDIF) ;
    b. http://decrochage.celso.co (cours conçu par une étudiante de la promotion 2013–14 du MARDIF) ;

c. http://webu2.upmf-grenoble.fr/sciedu/pdessus/ue28.html

d. Lecture attentive de la documentation en français (Le Clainche, 2010), accessible à http://formatice.org/telechargement/exe/guide_utilisateur.pdf. Les récentes version d'eXeLearning contiennent une aide en ligne en anglais claire et exhaustive (menu Aide>Manuel eXe).

D'autres documentations sont disponibles :

e. EdutechWiki (2011) à http://edutechwiki.unige.ch/fr/EXe.

f. Des diapositives de présentation à http://jil.quillevere.free.fr/CARIF_LR_14112006/crer_avec_exelearning.html

g. Notez que le manuel de référence (en anglais) est accessible à http://wikieducator.org/Online_manual.

3. En quelques phrases, écrire ce que l'on a compris/pas compris sur le logiciel. S'en servir pour la classe virtuelle.

4. Installation du logiciel, à partir de la page http://exelearning.net/downloads/. Au 12 février 2015 c'est la v 2.01 la version stable la plus récente. Dans le cas où votre ordinateur serait ancien et que vous rencontrez des problèmes de stabilité, le site http://exelearning.org répertorie de plus anciennes versions du logiciel. Notez qu'à partir de la v. 2.0, de nouveaux outils pédagogiques (Pendu, Memory, Tri d'éléments) sont disponibles en version ß. Ne pas hésiter à les tester s'ils correspondent à des activités que vous voulez faire faire à vos apprenants.

5. Il est présupposé deux éléments importants (voir aussi formulaire à signer ci-après) :

   a. Sauf avis contraire de l'étudiant, mentionné clairement à l'enseignant par courriel le plus tôt possible, le cours produit sera placé sous *licence Creative Commons BY-NC-SA* (plus d'informations à http://creativecommons.org/licenses/by-nc-nd/2.0/fr/). Le logiciel *eXe Learning* permet d'ailleurs d'ajouter automatiquement une telle licence sur toutes les pages produites.

   b. La production réalisée devra être exempte de contenu plagié. Il appartient donc à l'étudiant de référencer scrupuleusement le contenu provenant d'auteurs autre que lui ou elle-même. Ce contenu sera nécessairement : – court (proportionnellement au contenu écrit par l'étudiant) ; – mentionné entre guillemets ; – suivi ou précédé de sa référence (article, ouvrage). Tout manquement à ce point pourra faire l'objet de sanctions par l'instance universitaire appropriée. La consultation du site http://www.integrite.umontreal.ca/ pourra apporter d'utiles précisions.

B. Conception du cours

1. Choisir un domaine dans lequel vous êtes expert-e et pour lequel vous voulez transmettre un contenu.

2. Choisir un niveau de classe et définissez quelques objectifs d'apprentissage (ce que les personnes qui suivront ce cours devront être capables de faire à l'issue du cours).

3. Choisir un contenu à apprendre à propos du domaine choisi. Rédiger cinq pages A4 de cours (soit entre 17 000 et 20 000 caractères, espaces compris) hiérarchisées à deux niveaux. Veiller à ce que le contenu soit également réparti dans les différentes sections du cours.

4. Illustrer ce cours par des images, des tableaux, des vidéos, dont la référence sera mentionnée dans le corps du texte du cours.
5. Documenter les intentions pédagogiques du cours et son public. Chacun des points suivants doit faire l'objet d'une sous-section dans le cours *eXe Learning*, d'après Andrade *et al*. (2008) :
   a. Qu'est-ce qui doit être appris ? (détailler le contenu du cours, les principaux concepts et habiletés à l'œuvre).
   b. Pourquoi cela doit-il être appris ? (expliquer le contexte d'apprentissage et les possibles transferts du contenu du cours à d'autres contextes, ainsi que les contextes dans lesquels le contenu pourrait être utile).
   c. Comment cela doit-il être appris ? (détailler les méthodes pédagogiques, les types d'exercices utilisés dans le cours).
   d. Quand cela doit-il être appris ? (détailler les pré-requis, les éléments nécessaires avant de pouvoir apprendre ce contenu).
   e. Où cela doit-il être appris ? (détailler les éléments de contexte : p. ex., cours à distance, etc. relié au contenu à apprendre, mais aussi dans quel *curriculum* ce contenu s'insère.)
   f. Qui doit apprendre ce contenu ? (détailler les principales caractéristiques des apprenants : âge, compétences, niveau d'enseignement, etc.).
6. Concevoir au moins 5 questions interactives (de votre choix) qui utilisent les fonctionnalités d'*eXe Learning*. Parmi ces 5, l'une d'entre elles doit être une question Assertion-raison. Ces questions doivent être correctement insérées dans le cours et permettre une meilleure compréhension de ce dernier par le lecteur. Elles doivent, d'autre part, se conformer aux différentes prescriptions contenues dans ce cours.
7. Si le contenu de toutes les sections précédentes peuvent tout à fait être conçues et écrites sur un simple logiciel de traitement de textes, il est maintenant nécessaire de porter ce contenu, *dans son intégralité*, dans *eXe Learning*. Il est pour cela nécessaire d'utiliser la fonctionnalité « plan » de ce logiciel pour ajouter autant de pages qu'il est nécessaire (concevoir des pages de cours seul ou bien des pages mixtes cours puis QCM), puis les peupler du contenu (par un simple copier-coller) du cours et créer les QCM un à un (là aussi, recourir au copier-coller).
8. Il faut ensuite placer le cours exporté en HTML sur un serveur web. Les étudiants du MARDIF trouveront des renseignements sur ce point dans la 3e partie de ce document.

Présentation rapide d'Oppia

Voici les liens vers les principales informations sur *Oppia.*

Site général : https://www.oppia.org/

Utiliser des Explorations : https://www.oppia.org/learn

Modifier une Exploration ou en créer une : https://www.oppia.org/contribute

Documentation : https://code.google.com/p/oppia/wiki/Documentation

Code : https://code.google.com/p/oppia/

Introduction

Oppia (https://www.oppia.org/) est un système conçu initialement par des employés de *Google* (sans être un produit *Google*, comme cela est indiqué ici https://www.oppia.org/about#/what-is-this-website) permettant de concevoir assez aisément des activités d'apprentissage en ligne. Il est *open source* (le code peut être récupéré ici : https://code.google.com/p/oppia/) et pourra donc être sujet à de nombreuses améliorations.

Un enseignant, dans *Oppia*, peut générer des parcours (appelés Explorations) dont les étapes (appelées Nœuds) représentent différentes interactions dialogiques, enchaînements de questions et rétroactions qui peuvent être personnalisées, comme si un enseignant posait une série de questions à un apprenant en analysant ses réponses pour le mener à une prochaine question. L'originalité d'*Oppia* est que l'ensemble de ces interactions sont affichées sur une seule page, pour avoir un historique complet des interactions.

Ces Explorations peuvent être générées sur un serveur en ligne à partir d'exemples déjà réalisés (https://www.oppia.org/learn) qui peuvent être copiés (clonés) et modifiés, puis publiées, ou bien directement réalisées sur son propre serveur (cela nécessite une installation plutôt complexe, toutefois).

Comparativement à d'autres outils de génération de parcours d'apprentissage (comme *eXe Learning*, http://exelearning.net/), *Oppia* est parfois plus performant (dans la finesse des rétroactions, l'utilisation de variables, ou la production de traces de réponses ou parcours), parfois moins (les activités sont peu variées, il n'y a pas la possibilité de créer un test de closure, par exemple).

Découverte des fonctionnalités d'Oppia

Ce document n'est pas un mode d'emploi complet d'*Oppia*. Il présente seulement quelques premières pistes d'utilisation. Voici une suite de conseils pour comprendre ses fonctions de base.

1. Visionner la vidéo présentant *Oppia* (en anglais) : https://code.google.com/p/oppia/
2. Naviguer dans différentes Explorations, classées par matière (https://www.oppia.org/learn) afin de comprendre les principales fonctions d'*Oppia*. Attention, l'interaction se fait en anglais la plupart du temps.
3. Une fois ce premier tour réalisé, vous pouvez vous pencher sur l'Exploration « Villes de France » (Catégorie Geography), qui contient l'ensemble des Nœuds décrits plus bas. Vous pouvez l'éditer (mode contribute) ou la cloner, pour comprendre en détail comment elle fonctionne.
4. Trouver ensuite une Exploration dont la structure est assez proche de ce que vous comptez faire (« Villes de France » ou une autre). Passez en mode création (https://www.oppia.org/contribute) et clonez-la (il sera nécessaire de vous créer un compte *Google* pour cela). Cela crée une copie privée de l'Exploration que vous pouvez modifier pour l'adapter à votre contenu et besoins. Vous pourrez ultérieurement la rendre publique après demande aux gestionnaires du site. Notez que la création d'une Exploration (sans clonage) permet de la rendre publique sans demande préalable.

Une première Exploration

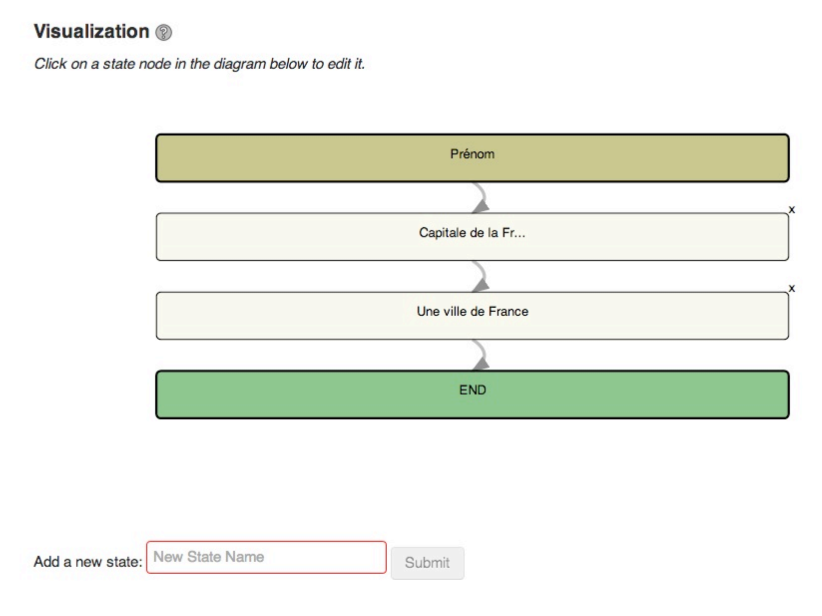
Toute Exploration est composée d'une suite de « Nœuds d'état » (*State Nodes*), désormais Nœuds, dont on peut définir finement l'enchaînement et les interactions. Tout Nœud comprend les éléments suivants :

— les noms des variables (*parameters*) qui sont utilisés dans le Nœud, pouvant provenir de Nœuds précédents (voir le passage du prénom de l'apprenant dans l'Exploration décrite ci-dessous).
— *un contenu* (*content*) : texte, image ou vidéo informative, ou encore question ;
— *une interaction* (*interaction*) : zone d'entrée d'un texte, nombre, questionnaire à choix multiple
— *des règles,* affichant une rétroaction en fonction des données de la phase précédente, et orientant ensuite l'apprenant à un autre Nœud.

Chaque étape affichée par un Nœud s'affiche à la suite des précédentes étapes, provenant des Nœuds précédents, sans changement d'écran, un peu comme si l'apprenant suivait une conversation. Ce document présente une Exploration, tout à fait minimale, mais permettant de comprendre le fonctionnement d'*Oppia*, commence par demander le prénom de l'apprenant. Ensuite, une question sur la capitale de la France est donnée, puis une reconnaissance de ville française d'après une photo.

La Fig. 2 ci-dessous représente l'enchaînement du parcours minimal d'une exploration. Les Nœuds initial et final sont plus foncés. Les flèches mentionnant l'enchaînement possible entre Nœuds (parcours dans l'Exploration) sont créées automatiquement en fonction des choix. Dans l'exemple, l'apprenant commence par donner son prénom. Ensuite, une question sur la capitale de la France, puis sur une ville, sont données. Il est possible d'ajouter des Nœuds en spécifiant un nouveau nom de Nœud dans le champ du bas. Attention, il n'est pas possible de copier des Nœuds, donc il est important de réfléchir à l'ordre des Nœuds à l'avance, et notamment du premier (qui ne peut être changé de place une fois déterminé).

Figure 2 — Le parcours d'une Exploration minimale, les « Villes de France ».



Il faut éditer le Nœud d'état en cliquant dessus. La fenêtre ci-dessous apparaît (Fig. 3), elle permet de régler les paramètres suivants (qui sont tous éditables en cliquant sur le bouton à droite de chaque paramètre) :

— les changements de paramètres (variables) (zone *parameter change*), permet de spécifier des variables qui vont stocker des informations sur le parcours (p. ex., le nom de l'apprenant, certaines de ses réponses à traiter ultérieurement, les nœuds par lesquels il est passé, etc.). Par défaut, il y a la variable {{answer}} (réponse), qui contient le texte de la réponse en cours (cela peut être le rang de l'item du QCM choisi). Les paramètres peuvent être utilisés tout au long de l'Exploration si l'on

spécifie bien les changements de paramètres des Nœuds qui en utilisent, et sont mentionnables dans tout champ texte inséré dans des doubles accolades. Les noms de paramètres sont sensibles à la casse et ne peuvent contenir que des lettres et des chiffres.

— le *contenu* (zone *content*), c'est-à-dire la question ou l'information initiale. Ce contenu peut être textuel, imagé, en tableau, comprendre une formule mathématique (formule *LaTeX*), ou une vidéo. Des onglets peuvent également être affichés (pour indiquer une démarche en plusieurs étapes ou bien des indices successifs). Dans la Fig. 3, c'est une simple phrase, la question initiale « Quel est votre prénom ? ». Dans la Fig. 5, c'est une phrase accompagnée d'une photo.

— l'*interaction*, c'est-à-dire les différents items de choix (ici, les différentes villes), mais, là aussi, il est possible d'insérer de nombreux éléments de contenu (les mêmes que ceux de la rubrique précédente).

— les *règles* (zone « *when readers submit* ») à appliquer, en fonction du choix de l'utilisateur. Chaque règle est de type "si-alors" et est appliquée dans l'ordre d'affichage. Elle oriente enfin l'apprenant vers un autre Nœud. Dans l'exemple de la Fig. 4, la première règle est la suivante : « *Si* la réponse est égale à Berlin, *alors* afficher la rétroaction textuelle "Désolé, {{nom}}, Berlin est bien une capitale, mais de l'Allemagne, réessayez" » et demande à l'apprenant de resoumettre une réponse (boucle sur le Nœud « Capitale de la France »).

Attention, il est nécessaire, une fois l'édition d'un paramètre faite, de cliquer d'abord sur « *Use this interaction* » (ou « *Use this content* »), de cliquer sur le bouton — en haut à droite de la fenêtre, puis de cliquer sur « Save Changes » (bouton en haut à droite) dans la fenêtre principale pour que les changements soient pris en compte. La suite de cette section décrit pas à pas la création des Nœuds.

Etape 0 — Créer une Exploration

Une nouvelle exploration se crée de la manière suivante :

— Si vous n'en avez pas déjà un, créez-vous un compte Google et connectez-vous au site d'*Oppia*.

— De la page https://www.oppia.org/contribute, cliquez sur « Create New Exploration ».

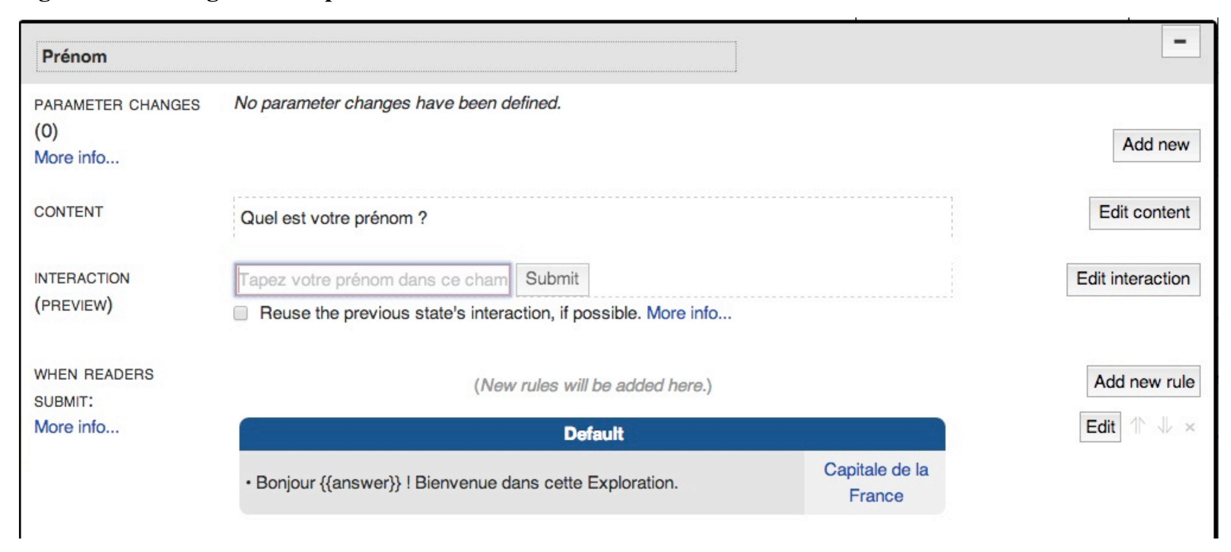
— Mentionner son titre et sa catégorie dans la fenêtre qui s'ouvre.

— Une page s'ouvre, avec, dans la section « Visualization », 2 Nœuds créés par défaut : celui du début et celui de fin.

Etape 1 — Enregistrer le prénom de l'apprenant

Cliquez sur le Nœud de début et nommez-le « Prénom ». La première étape est de créer un Nœud qui demande le prénom de l'apprenant pour pouvoir s'en servir dans les Nœuds suivants (voir Fig. 3). Si l'on devait juste afficher le prénom dans l'interaction du Nœud courant, l'usage du paramètre par défaut {{answer}} suffirait. La fenêtre du Nœud doit être configurée comme suit :

Figure 3 — Enregistrer un prénom.



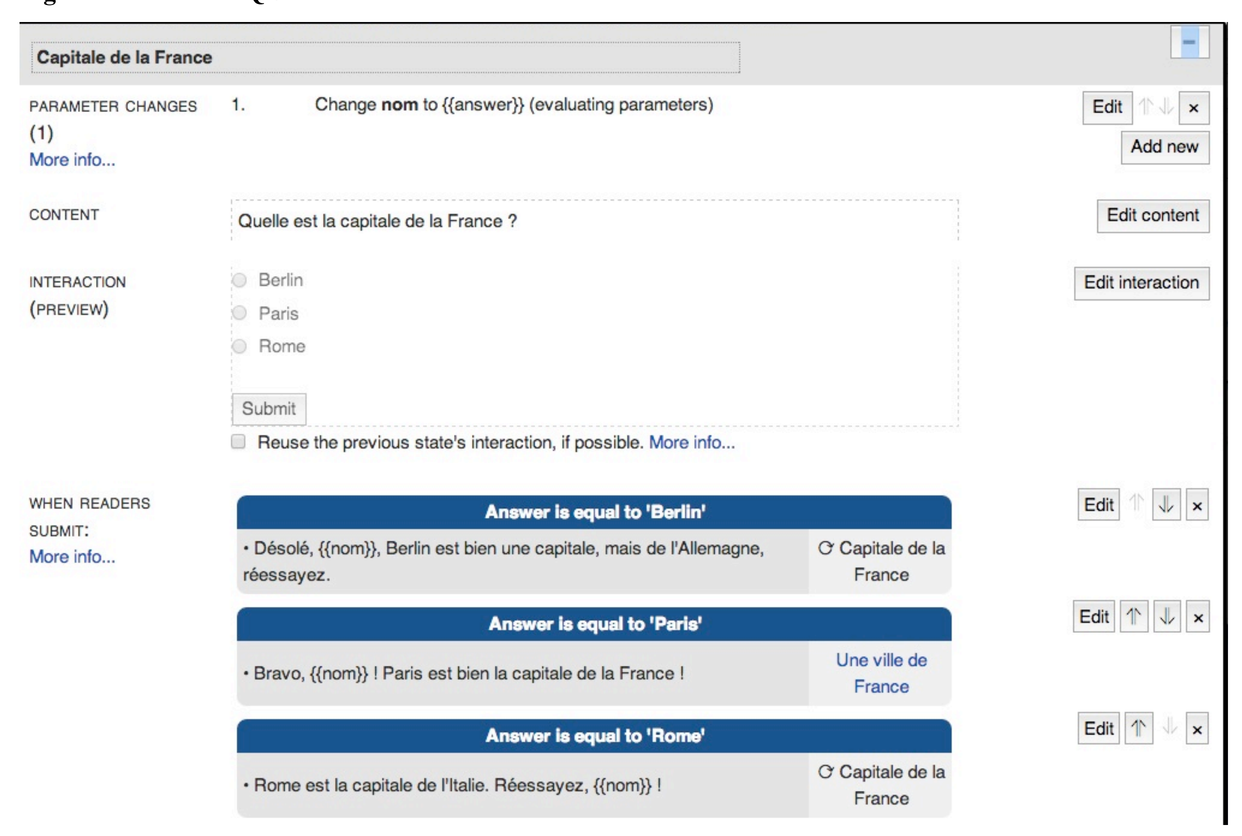
— cliquer sur « Edit Content » pour taper la question demandant la saisie du prénom.
— cliquer sur « Edit Interaction » pour choisir l'interaction « Text Input », qui peut être paramétrable, par exemple en modifiant la consigne « Tapez votre prénom dans ce champ ». La taille du champ de saisie (ligne-colonnes) est également spécifiable.
— cliquer sur « Add new rule » pour souhaiter la bienvenue à l'apprenant. Vous notez qu'il est nécessaire ensuite de l'orienter vers le Nœud suivant, « Capitale de la France ». Le paramètre {{answer}} est utilisé ici pour afficher le nom de l'apprenant car on reste à l'intérieur du Nœud. Vous verrez par la suite comment utiliser cette donnée dans un autre Nœud.

Etape 2. Un QCM

Une fois ce premier Nœud réalisé. Passons au deuxième, qui intègre un QCM sur la capitale de la France. Créez un deuxième Nœud intitulé « Capitale de la France » et la fenêtre du Nœud doit être configurée comme ci-dessous (Fig. 4).

– Tout d'abord, il est nécessaire de faire passer la réponse du Nœud précédent dans une autre variable (puisque le paramètre {{answer}} va contenir la réponse au QCM et donc effacer le prénom). Le passage de paramètres se fait en ajoutant un paramètre de manière à ce que "Change nom to {{answer}}" et en cochant obligatoirement la case "evaluating parameters". Cette fois, c'est le paramètre {{nom}} qui devra être utilisé pour afficher le prénom. Une fois ce changement de paramètre fait, il n'est plus nécessaire de le mentionner à nouveau dans d'autres Nœuds qui l'utilisent.
– Ensuite, il faut Editer la zone Contenu (*Edit content*) pour afficher la question « Quelle est la capitale de la France ? »
– Puis, Editer la zone Interaction (*Edit Interaction*), en choisissant l'option « Multiple Choice Input » et indiquant les 3 réponses possibles. Vous noterez qu'il est aussi possible de choisir des images, des vidéos, etc.
– Ensuite, il faut éditer les différentes réponses possibles dans les Règles (*When Readers Submit*), comme indiqué sur la Fig. 4.

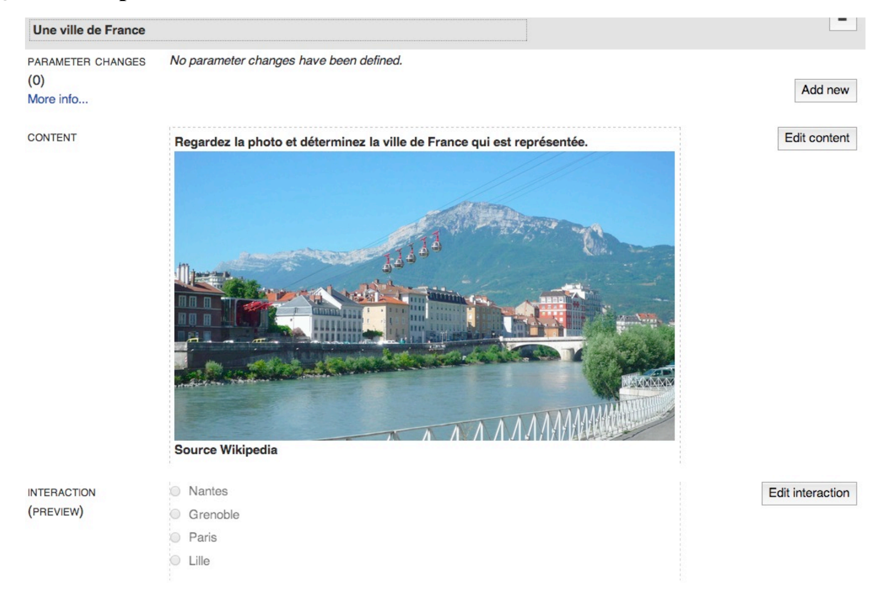
Figure 4 – Faire un QCM.



Etape 3. Un QCM avec photo

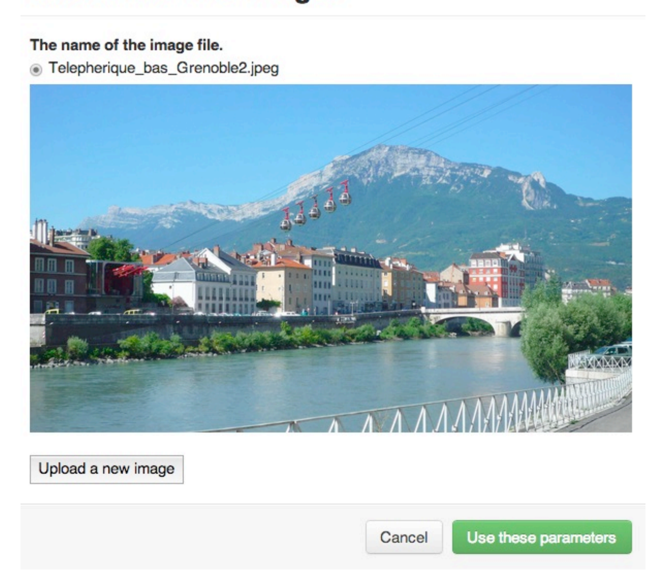
La troisième et dernière étape concerne aussi un QCM, mais cette fois avec une photo, insérée dans le champ « Contenu ». La Fig. 5 ci-dessous en décrit les principaux paramètres, qui varient peu de ceux de la Figure précédente.

Figure 5. Un QCM avec photo



Notez simplement que le passage de paramètres pour le prénom n'est plus nécessaire, car il a été fait au précédent Nœud. L'insertion de la figure, qui se fait assez aisément (voir Fig. 6). Il faut tout d'abord téléverser l'image sur le serveur (elle doit avoir l'extension jpeg), puis la sélectionner en cliquant sur le bouton.

Figure 6. Insérer une image dans un contenu.



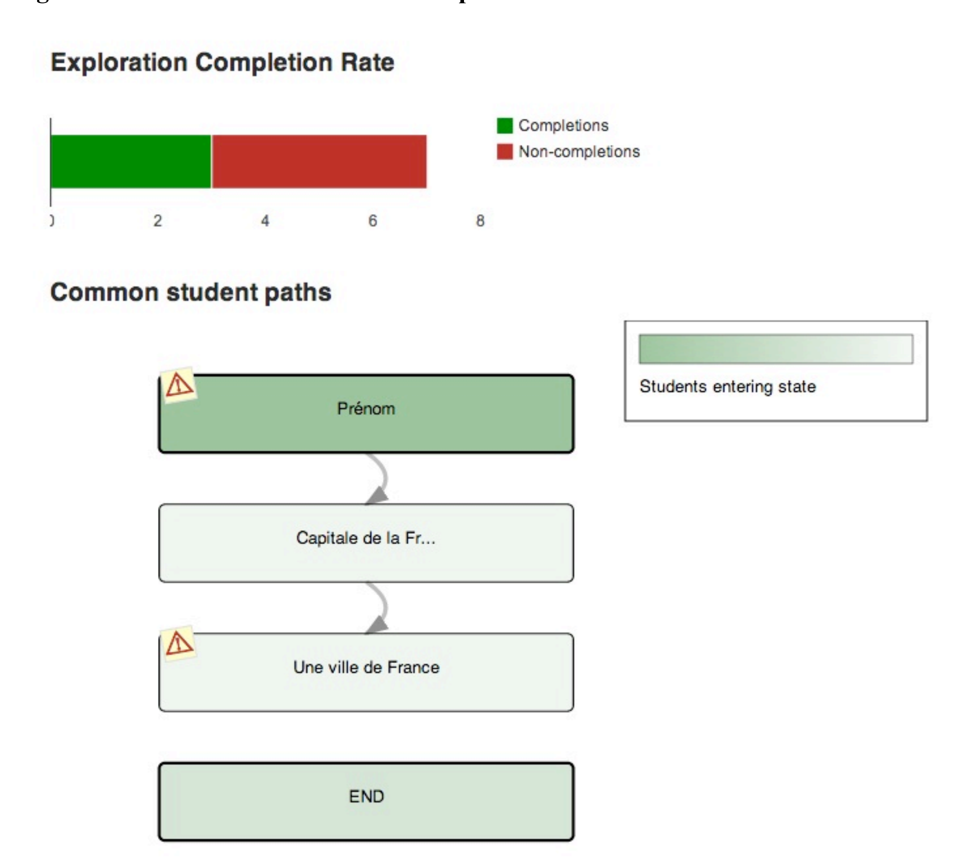
Jouer l'Exploration

Une fois l'Exploration conçue, il est aisé de la jouer pour être sûr qu'elle corresponde à ce qu'on veut. Ensuite, il est possible de la publier (bouton « Publish ») pour la rendre accessible à tous.

Fonctionnalités de traces

L'un des intérêts d'*Oppia* est qu'il récupère l'ensemble des parcours des apprenants (cliquer sur l'onglet « Exploration stats » (voir Fig. 7) pour avoir une idée du nombre d'apprenants ayant parcouru l'ensemble de l'Exploration. Un clic sur chaque nœud affiche les différentes réponses obtenues, ce qui permet d'avoir une idée de la difficulté des questions, et aussi d'élaborer une règle spécifique.

Figure 7— Visualisation des traces des parcours



Voir plus loin

Le site LDShake (http://ldshake3.upf.edu/), accessible sur inscription gratuite, permet de créer et partager avec une communauté d'enseignants des documents de cours créés sous *eXe Learning*.

10. Analyse *a posteriori* des items du questionnaire

Nous avons vu, à la Section 8, quelques principes et recommandations pour concevoir des QCM efficaces. Ces principes interviennent *a priori*, pendant que l'enseignant construit son QCM et avant que ses élèves le passent. Il y a un autre moyen, *a posteriori*, d'améliorer la qualité de chaque item, que nous allons voir ici.

Crisp et Palmer (2007) ont montré que les enseignants (du supérieur, dans leur article) étaient insuffisamment formés aux techniques d'analyse statistique des QCM qu'ils pouvaient faire passer à leurs élèves (ils ne calculent en effet, en majorité, que la moyenne et la médiane des résultats), et qu'une simple batterie de tests statistiques pouvait leur permettre, notamment, d'améliorer la qualité de ces derniers. Comme le signale Haladyna (2004), avec un certain humour, il existe des régularités dans les réponses des apprenants aux QCM, régularités que peuvent donc capturer des tests : « […] les élèves avec de hauts niveaux de connaissance ou habileté tendent à choisir la bonne réponse et les élèves avec de bas niveaux de connaissance ou habileté tendent à choisir la mauvaise » (*id*., p. 203).

Cette section vise à donner quelques indications pour l'analyse *a posteriori* des résultats du questionnaire. Une fois le questionnaire passé auprès d'apprenants, il est nécessaire d'analyser de plus près les réponses afin de :

- se faire une idée du niveau de connaissances des personnes l'ayant passé, en analysant apprenant par apprenant et question par question leur réussite ou échec ;
- redistribuer l'ordre des questions, de manière, par exemple, à placer les questions faciles au début et les questions plus difficiles vers la fin ;
- voir si certains items sont vraiment utiles et considérer leur suppression,
    - soit parce qu'ils sont trop ou pas assez *difficiles* (proportion de personnes ayant sélectionné la bonne réponse trop élevée ou trop basse), ce qui peut s'analyser à partir de la proportion de personnes ayant sélectionné telle ou telle réponse d'une question ;
    - soit parce qu'ils sont pas assez *discriminants* (c'est-à-dire qu'ils ne permettent pas bien de distinguer les étudiants performants dans le domaine du QCM de ceux moins performants, ce qui se passe lorsque tous deux répondent correctement à l'item, quel que soit leur niveau), ce qui peut s'analyser en calculant la corrélation entre le score à un item et le score total du questionnaire, par élève.
- voir si certains distracteurs ne jouent pas leur rôle, parce qu'ils sont pas assez – ou même trop – choisis.

Il faut toutefois faire remarquer au lecteur les points suivants. Tout d'abord, ces processus d'analyse ne remplacent pas une analyse soignée par l'enseignant de la validité de chaque question (c.-a.-d., se demander si la réponse à la question met bien en œuvre la compétence et la connaissance voulue par lui). Ils peuvent toutefois alerter l'enseignant sur des items pouvant poser des problèmes.

De plus, ce type d'analyse ne peut se faire que pour un nombre suffisamment important d'élèves. Haladyna mentionne le nombre minimal de 20, tout en considérant qu'un « petit » effectif en comprend moins de 200. Tout calcul réalisé sur des échantillons de moins de 20 élèves est donc à prendre avec la plus grande précaution.

Ensuite, les indices dont nous préconisons le calcul ci-après sont *dépendants du niveau des élèves ou étudiants* qui passent le test. Cela signifie qu'ils ne permettent pas un calcul (resp.

de la difficulté et du pouvoir discriminant) des items dans l'absolu, mais lié au niveau des élèves y répondant. Il est donc tout à fait possible qu'un item détecté pour une passation et un groupe d'élèves comme difficile soit analysé comme plus facile avec un autre groupe et *vice versa*. Là aussi, le jugement de l'enseignant prévaut.

Etudions plus en détail ces indices. Pour les calculer, un simple tableur convient (*LibreOffice*, *OpenOffice* ou *MS Excel*) peut être utilisé pour les calculs ci-dessous et les formules sont à la portée de tout utilisateur averti. Le lecteur trouvera plus de détails, dans une sous-section à venir, à propos d'un fichier tableur que nous avons créé à cet effet. La feuille *CITAS* (*Classical Item and Test Analysis Spreadsheet*), disponible gratuitement pour *MS Excel* et *OpenOffice*, à http://www.assess.com/xcart/product.php?productid=407 peut également être utilisée et contient des indices plus avancés. La majorité des logiciels de statistique avancés, comme *R*, *SAS* ou *SPSS* peuvent également calculer ces indices.

Analyse d'items : la difficulté/facilité des questions (P)

Le premier indice que nous calculons est lié à la difficulté/facilité de chacun des items du QCM. Il est tout à fait nécessaire que le questionnaire comprenne quelques questions plutôt faciles, de manière à rassurer l'élève ou l'étudiant. En revanche, une trop grande proportion de questions faciles amène à un « effet-plafond » (et un effet-plancher pour des questions trop difficiles) et donc ne représente pas une tâche suffisamment consistante et cognitivement engageante pour la plupart des élèves.

L'indice P_i (variant entre 0 et 1) est la proportion d'élèves ayant répondu juste à un item *i*. Un item est d'autant plus facile qu'il a un *P* élevé. De plus, il faudrait que la majorité des items aient un *P* se distribuant autour de 0,5. Comment apprécier cette valeur ? McAlpine (2002b) considère que les réponses trop difficiles ($P < 0,15$, soit lorsque moins de 15 % des participants ont répondu juste à la question donnée) ou trop faciles ($P > 0,85$, soit lorsque au moins 85 % y ont répondu juste) peuvent être écartées du QCM. Les premières sont donc « impossibles », les secondes « triviales » et donc ni l'une ni l'autre ne permettent la mesure efficace des compétences des personnes passant le QCM. Il faut noter que, selon les auteurs, ces valeurs varient : 80 % et 25 % (Bodner, 1980) et que, d'autre part, des raisons pédagogiques peuvent amener à conserver des questions au-delà ou en-deçà de ces seuils.

L'indice *P*, en lui-même, n'est pas suffisant pour juger de la qualité des items, et donc pour les accepter ou rejeter définitivement. Il est nécessaire pour cela de s'intéresser à leur pouvoir discriminant.

Analyse d'items : le pouvoir discriminant des questions (Indice de discrimination ou ID)

Le deuxième indice que nous calculons concerne le pouvoir discriminant de chaque item. Le but d'un questionnaire *peut être* de mesurer une compétence/connaissance donnée, et donc de pouvoir discriminer les élèves en fonction de cette dernière. Bien évidemment, il est possible de considérer exactement le contraire : que le questionnaire n'est pas fait pour discriminer ou sélectionner, mais pour simplement certifier des compétences (test critérié). Dans ce dernier cas, l'*ID* peut tout de même être utile pour détecter les items mal formulés ou ayant trait à une dimension de connaissance/compétence différente de celle mesurée par le test. Il sera aussi possible de calculer l'indice de Brennan, qui est utile pour des situations d'enseignement formatives.

Nous l'avons déjà évoqué dans l'introduction de cette section : l'indice de discrimination (*ID*) est une variante du coefficient de corrélation de Pearson et est nommé *r point-bisérial* (r_{pbis}), à utiliser lorsque l'une des deux variables est dichotomique (en effet, les scores à un item sont soit 0 soit 1). Il varie, comme tout coefficient de corrélation, entre –1 et 1, et correspond au

coefficient de corrélation entre un item et le score total au test. Il est ensuite important de voir dans quelle mesure une question donnée contribue, par elle-même, à sélectionner de « bons » élèves ou de « moins bons ». La valeur du r_{pbis} d'un item donné peut être :

– *positive* et élevée, ce qui signifie que les élèves réussissant bien au test réussissent également bien à l'item concerné.
– *négative*, ce qui signifie, soit que ce sont les élèves peu performants en général (sur l'ensemble du questionnaire) qui réussissent bien à l'item, ce qui peut laisser penser que cet item est un faible prédicteur de la connaissance ou compétence mesurée dans le test, et donc qu'il faut considérer sa suppression. Soit encore que les élèves performants en général réussissent moins bien à cet item : il convient dans ce cas de se demander pourquoi (il faut au passage noter qu'un r_{pbis} élevé et négatif est rarement rencontré, et témoigne très souvent d'une erreur de saisie des données).
– nulle ou faible (en valeur absolue inférieure à 0,20), ce qui signifie que le lien entre la réussite à l'item et la réussite au test est nul ou faible. L'enseignant peut considérer sa suppression du test.

Trois points sont à noter.

– Un item discriminant n'est pas nécessairement difficile. À l'inverse, un item problématique (donc à rejeter ou reformuler) a presque toujours un *ID* faible.
– Si une question a une corrélation item-test faible (en valeur absolue inférieure à 0,20), cela peut signifier qu'elle mesure une compétence/connaissance qui n'appartient pas au même domaine que celui du test complet (un pouvoir discriminant est satisfaisant si la corrélation est supérieure à 0,40). Même si nous devons le déconseiller dans le cadre de ce cours, si le test est conçu pour être multidimensionnel (mesure plusieurs types de compétences/connaissances), la corrélation est à réaliser entre la question et le sous-ensemble de questions liées à une compétence particulière.
– Enfin, un test a une bonne cohérence interne (*i.e.*, il est composé d'items qui concourent à mesure un type de connaissance/compétence) si ses items ont des *ID* élevés (plutôt proches de 1). Il existe un autre indice rendant compte de cette cohérence en une seule valeur (*alpha* de Cronbach, ou *Kuder-Richardson*), mais son étude sort du propos de ce document.

Pour finir, il est utile d'avoir cette mesure de discrimination entre bons faibles élèves seulement si on se place dans une optique différenciatrice (ou sommative). Cela n'est pas du tout l'objectif d'un enseignant qui se place dans une démarche formative, ou de pédagogie de la maîtrise. Dans ce cas, il importe moins de déterminer les bons/faibles que de déterminer ceux qui maîtrisent/ne maîtrisent pas un contenu donné (au-dessus d'un seuil fixé par l'enseignant). Dans ce cas, on s'intéresse à calculer un « indice de discrimination au seuil de maîtrise » ((Laveault & Grégoire, 2002, pp. 222–223), ou indice de Brennan (variant entre –1 et 1).

Il convient tout d'abord de fixer le seuil de maîtrise. Bloom (Bloom, 1988), dans la pédagogie de la maîtrise, stipulait qu'un élève maîtrisait un contenu s'il réussissait 80 % des items). Il est donc aisé de faire deux groupes d'élèves, l'un contenant tous les élèves réussissant ces 80 % du test, l'autre groupe contenant tous les autres élèves.

L'indice de Brennan (*B*) est simplement la différence entre deux indices de difficulté d'un item : celui pour ceux ayant atteint le seuil de maîtrise dans le test entier (groupe « maîtrise ») auquel on soustrait ceux ne l'ayant pas atteint (groupe non-maîtrise). Un indice positif montre le pourcentage d'élèves du groupe « maîtrise » a mieux réussi l'item que dans le groupe

« non-maîtrise ». Si la valeur B est négative, cela veut dire que c'est le groupe « non-maîtrise » qui réussit mieux l'item. Ce dernier ne discrimine donc pas les élèves dans le bon sens et on peut considérer à le supprimer. Attention, il convient d'utiliser des notes ou des critères différents de ceux utilisés pour calculer l'indice, afin d'éviter tout problème de circularité.

Le Tableau 6 ci-dessous et les calculs plus bas indiquent comment calculer B. Les colonnes du Tableau indiquent les élèves par groupe de maîtrise (14 élèves sont dans le groupe « non-maîtrise et parmi eux seuls 5 ont réussi l'item ; 20 élèves sont dans le groupe « maîtrise » et parmi eux 16 ont réussi l'item).

Tableau 6 — Organisation des données pour le calcul de l'indice *B* de Brennan (Laveault & Grégoire, 2002, p. 222).

	Réussi $a + b$	$a = 5$	$b = 16$
Item	Echoué $c + d$	$c = 9$	$d = 4$
		Non-maîtrise : $a + c$	Maîtrise : $b + d$

Ensuite, il est aisé de calculer les indices de difficulté de l'item pour ces deux groupes, resp. P_m et P_{nm}

$P_m = b/(b+d) = 0{,}8$

$P_{nm} = a/(a+c) = 0{,}36$

$B = P_m - P_{nm} = 0{,}8 - 0{,}36 = 0{,}44$

Cela signifie que si on compare la réussite des élèves du groupe « maîtrise » à celle du groupe « non-maîtrise », il y en a 44 % de plus qui réussissent l'item dans le premier groupe. On peut considérer que cet item est à conserver. Un onglet du tableur Excel dont l'URL figure dans la section suivante permet de calculer B.

Le rôle des réponses distractrices

La troisième et dernière procédure va s'intéresser, item par item, aux réponses distractrices. Tout item de QCM contient une bonne réponse et des réponses distractrices. Il convient aussi de vérifier si les différents distracteurs ont bien joué leur rôle (pour un distracteur donné, si personne ne l'a sélectionné il ne l'est pas, si tout le monde l'a sélectionné, il l'est sans doute trop et les étudiants ont été massivement induits en erreur). Johnstone (2003) considère qu'un distracteur remplit son rôle si environ 5 % *au moins* des personnes le choisissent. La valeur de choix maximale dépend bien sûr du nombre de réponses, mais il est possible, en affichant une représentation graphique de tous les choix question par question, de déterminer les réponses distractrices qui ont été massivement évitées : il faudra sans doute les reformuler.

Le moyen de déterminer si les réponses distractrices remplissent bien leur rôle est simple : il suffit de représenter graphiquement une distribution par item et par réponse, et de vérifier visuellement la distribution des choix. La figure ci-dessous en donne un exemple.

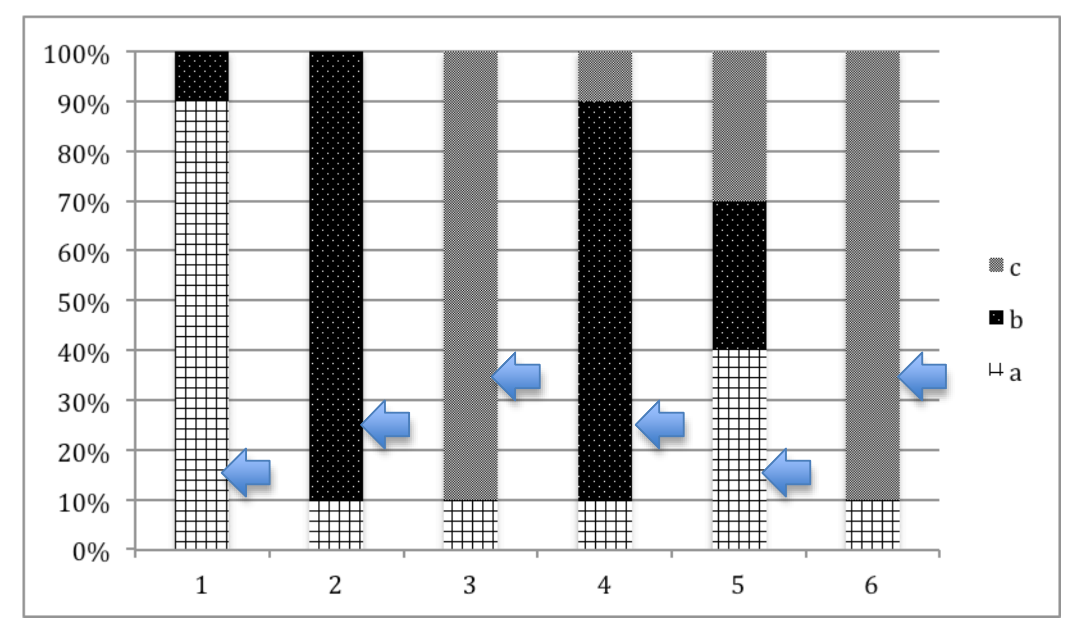


Figure 8 – Pourcentage des réponses des élèves par item (1–6) et par réponse (a–c). Les flèches indiquent les bonnes réponses.

Une analyse de ce type de graphique peut nous permettre de tirer les conclusions suivantes. Tout d'abord, il est possible de déterminer si la place de la bonne réponse se distribue uniformément entre les différents items (ici il y a uniformité). Ensuite, il est possible de détecter les distracteurs jamais choisis (p. ex., *c* pour l'item 2) et de le reformuler de manière à le rendre plus proche du bon résultat, tout en étant faux. Enfin, comme le suggèrent Crisp et Palmer (2007), lorsque les items sont de difficulté croissante, il est ainsi possible de vérifier qu'au début, les élèves choisissent plus aisément la bonne réponse qu'à la fin du questionnaire (ici, ce sont les questions 4 et 5 qui ont les distracteurs les plus attractifs).

Une stratégie d'analyse des items

L'examen attentif et parallèle des deux derniers indices (difficulté et discrimination) permet d'analyser finement la qualité des items. Il est facile de régler les cas où les indices *P* et *ID* d'un item sont conjointement bas (item très difficile et peu discriminant, donc à rejeter) ou conjointement élevés (item très facile et très discriminant, donc à conserver, mais en gardant également des items plus difficiles). Evoquons les autres cas.

Lorsqu'un item a un *ID* élevé et un *P* bas, cela indique que seuls les meilleurs élèves ont réussi l'item. Cet item très sélectif pourra être conservé, là aussi en le mixant avec des items plus faciles. Lorsqu'un item a un *ID* faible et un *P* élevé, il faut en analyser la raison élève par élève : il est très probable qu'un ou plusieurs bons élèves (*i.e.*, ayant bien répondu à des items plus difficiles que celui-ci) n'aient pas compris la consigne de l'item. Il est possible que cet item doive être reformulé ; mais aussi qu'il mesure des connaissances ou compétences non reliées aux autres.

Utiliser le fichier tableur pour calculer les différents indices

Le fichier tableur disponible à http://webu2.upmf-grenoble.fr/sciedu/pdessus/cours/tabQCM.xls et dans un format compatible *MS Excel*, *LibreOffice* et *OpenOffice* permet de calculer les différents indices ci-dessus. Il sera nécessaire d'adapter le tableau au nombre d'items et d'élèves du QCM passé, ce qui est aisé à faire en copiant/collant les cellules des différents tableaux.

Voici comment faire pour remplir les différents tableaux. Commençons par la feuille « Analyse distracteurs ». Soit un QCM à 6 items, passé à 10 élèves, qui ont chacun 3 réponses possibles (*a*, *b* ou *c*). Dans le cas où le QCM est passé avec plus d'élèves, comporte plus d'items ou de réponses, il est aisé de créer des lignes ou colonnes supplémentaires en en insérant, voire en en supprimant dans le cas contraire. L'enseignant aura ensuite à remplir le tableau comme celui de la Figure 3, qui montre par exemple que la bonne réponse de la question 1 est le choix *a*, et que l'élève *A* a donc trouvé la bonne réponse de cet item.

	A	B	C	D	E	F	G
1	**Elèves/Items**	**1**	**2**	**3**	**4**	**5**	**6**
2	**Bonne réponse**	**a**	**b**	**c**	**b**	**a**	**c**
3	**A**	a	b	c	c	b	c
4	**B**	a	b	c	b	c	a
5	**C**	a	b	c	b	a	c
6	**D**	a	b	c	a	a	c
7	**E**	b	b	c	b	b	c
8	**F**	a	b	c	b	a	c
9	**G**	a	b	c	b	a	c
10	**H**	a	b	c	b	b	c
11	**I**	a	a	a	b	c	c
12	**J**	a	b	c	b	c	c

Figure 9 – Tableau pour consigner les résultats du QCM par items et élèves.

Les colonnes référencent les différents items (questions) du QCM et les lignes les résultats de chaque élève. Il faut mentionner dans la ligne 2 la bonne réponse de chaque question et la réponse sélectionnée par chacun des 10 élèves (les lignes 3 à 12). Les autres tableaux se remplissent automatiquement et le graphique de la Figure 2 est mis à jour.

L'autre feuille de ce classeur, nommée « Calcul de *P* et *ID* », fonctionne de manière équivalente. Le calcul de ces indices est fait à partir du tableau items/élèves en 0/1 qui est généré dans la feuille précédente (en haut à droite) et il faut le recopier puisque aucun lien entre les feuilles n'a été créé (voir Figure 4, qui montre par exemple que l'élève A a réussi les items 1, 2, et 3 et échoué aux 4 et 5). Là encore, remplir ce tableau fait recalculer automatiquement toutes les autres valeurs. Au bas de la figure 4 sont détaillés, item par item, les indices *P* et *ID* (resp. sur les lignes grises et jaunes).

	A	B	C	D	E	F	G
1	**Elèves/Items**	**1**	**2**	**3**	**4**	**5**	**6**
2	**A**	1	1	1	0	0	1
3	**B**	1	1	1	1	0	0
4	**C**	1	1	1	1	1	1
5	**D**	1	1	1	0	1	1
6	**E**	0	1	1	1	0	1
7	**F**	1	1	1	1	1	1
8	**G**	1	1	1	1	1	1
9	**H**	1	1	1	1	0	1
10	**I**	1	0	0	1	0	1
11	**J**	1	1	1	1	0	1
12	score/item	9	9	9	8	4	9
13							
14		0,89	2,00	2,00	0,38	1,58	0,89
15		0,86	1,94	1,94	0,36	1,53	0,86
16	P	0,9	0,9	0,9	0,8	0,4	0,9
17	p faux	0,1	0,1	0,1	0,2	0,6	0,1
18		0,3	0,3	0,3	0,4	0,49	0,3
19	**r pbi (ID)**	**0,26**	**0,58**	**0,58**	**0,15**	**0,75**	**0,26**

Figure 10 – Tableau de données pour le calcul de *P* et *ID*.

Pour information, la formule du r_{pbis} est la suivante, où $\overline{X}_1$ est la moyenne (variable continue) de ceux qui ont eu 1 à la variable dichotomique ; $\overline{X}_0$ est la moyenne de ceux qui ont eu 0. S_y est l'écart type des scores de tous les participants. p est la proportion de participants ayant eu l'item juste.

$$r_{pbis} = \frac{\bar{X}_1 - \bar{X}_0}{S_y} \sqrt{\frac{p}{1-p}}$$

Voir plus loin

Il existe de nombreux autres indices, issus de la théorie des tests. Les modèles de réponse à l'item (ou IRT, *item response theory*) sont des modèles statistiques permettant de donner des indices « absolus », donc peu influencés par les caractéristiques de l'échantillon testé. Les ouvrages de Laveault et Grégoire (2002) et de Bernier (1985) sont de très bons ouvrages pour s'initier à ces questions.

Il existe aussi quelques logiciels, la plupart payants mais certains gratuits, qui aident les enseignants à calculer automatiquement les indices ci-dessus (voir section suivante). Ce tri fin, item par item, peut conduire au rejet d'un item particulier, mais aussi à la sélection d'items qui pourront être réutilisés lors d'examens suivants. La gestion régulière de QCM amène l'enseignant, s'il veut gagner du temps dans la sélection de questions, à les enregistrer dans une base de données. Des logiciels spécifiques existent (p. ex., http://fr.wikipedia.org/wiki/Logiciel_de_questionnaire_%C3%A0_choix_multiples), mais un simple logiciel de base de données peut très bien convenir.

11. Pour aller (encore un peu) plus loin

Le champ de la recherche sur les rétroactions automatiques est très actif. Les lectures suivantes pourront permettre au lecteur d'aller plus loin sur le sujet. Les ouvrages de Demaizière (1986) et Depover (1987), bien qu'anciens, permettront une bonne initiation à ce champ. Les ouvrages de Laveault et Grégoire (2002) et de Bernier (1985) initieront le lecteur sur la théorie (statistique) des tests. Tous les sites internet ci-dessous sont disponibles au 15 avril 2012, mais il est possible que des liens soient cassés à la date de consultation.

Pour réfléchir à des sujets de recherche

Voici quelques sujets de recherche que l'on peut mener dans le domaine du feedback automatique lié aux QCM.

1. Recherche historique sur les QCM en France (notamment dans les tests psychologiques).
2. Faire une revue technique des différents outils de création de QCM disponibles, et leurs avantages et inconvénients.
3. Comparaison de l'effet sur l'apprentissage de notions d'un QCM classique et d'un QCM Assertion-raison.
4. Effet de la passation d'un QCM (en tant qu'entraînement) sur l'apprentissage de notions (évaluées de manière sommative), dans la lignée des travaux de Roediger ou Bjork. Les évaluer en contexte de cours et non en laboratoire a été assez rarement réalisé.
5. Les effets du feedback sur la motivation des élèves (Georges & Pansu, 2011).

Pour en savoir plus sur la création de QCM (en français)

Le cours d'Osman (2011) est disponible à http://oer.avu.org/handle/123456789/228

Le document d'Audet (2011) est disponible à http://dev.refad.ca/evaluation_en_ligne.pdf

Pour en savoir plus sur la création de QCM (en anglais)

L'ouvrage de référence est celui d'Haladyna (2004). Quelques autres tutoriels de qualité :

Le *blog* de *Focal Works* :

Section 1 : http://focalworks.in/articles/creating-mcqs-part-i-ask-right-question

Section 2 : http://focalworks.in/articles/creating-mcqs-part-ii-make-right-choice

Section 3 : http://focalworks.in/articles/creating-mcqs-part-iii-alternative-types-multiple-choice-questions

Cheung et Bucat (2002), accessible à http://www3.fed.cuhk.edu.hk/chemistry/files/constructMC.pdf

Rodriguez (1997), accessible à http://www.edmeasurement.net/aera/papers/artandscience.pdf

Touchie (s.d.), accessible à http://www.mcc.ca/pdf/MCQ_Guidelines_e.pdf

Les guides de création d'items de l'UNESCO : http://www.sacmeq.org/research.htm#modules

Pour en savoir plus sur les logiciels

Il existe de très nombreux sites présentant, soit des logiciels autonomes ou des outils en ligne

générant des questionnaires. Le site http://webu2.upmf-grenoble.fr/sciedu/pdessus/sapea/logicielseval.html (de Flaugergues & Dessus, 2010) contient des informations techniques de base sur la problématique d'évaluer automatiquement des connaissances avec des outils informatisés.

Il existe aussi un très grand nombre d'outils générant des QCM. De Flaugergues et Dessus (2010) en listent certains, et en voici d'autres. Il est à noter que *Netquiz Pro*, d'un emploi un peu plus complexe qu'*eXe Learning*, permet de recueillir les résultats aux QCM de chaque apprenant, ce que ne permet pas ce dernier. Il convient de lire attentivement les conditions légales de leur utilisation avant toute exploitation, notamment éducative ou commerciale :

Concerto : http://www.psychometrics.cam.ac.uk/page/300/concerto-testing-platform.htm

Flax : http://flax.nzdl.org/greenstone3/flax

Google Forms : http://www.google.com/google-d-s/createforms.html

JClic : http://clic.xtec.cat/en/jclic/index.htm

LimeSurvey : http://www.limesurvey.org/fr/

LDShake : http://ldshake3.upf.edu/ (création et partage de documents *eXe Learning*)

MOS Solo : http://www.mindonsite.com/produits/mos-solo/ (Windows)

NetQuiz Pro : http://www.ccdmd.qc.ca/catalogue/netquiz-pro-4?section=block_6

QeDoc Quiz Maker : http://www.qedoc.org/en/index.php?title=Qedoc_Quiz_Maker

QuizFaber : http://quizfaber.free.fr/index.htm

RealQuest : http://www.neoptec.com/fr/produits/realquest/presentation.php

TAO: http://www.taotesting.com

Xerte : http://www.nottingham.ac.uk/xerte/index.aspx

De très nombreux outils *Web 2.0* sont également disponibles (attention à bien lire les conditions légales du site avant de créer son QCM), dont les suivants :

Blubbr (Quizzes avec YouTube) : http://www.blubbr.tv/index.php

Capira (insère des questions autocorrigées dans des vidéos) : http://capira42.appspot.com/

Fiftysneakers : http://fiftysneakers.com/

Flubaroo : http://www.flubaroo.com/

FunTrivia : http://www.funtrivia.com/

GoSoapBox : http://www.gosoapbox.com/

LearningApps : http://learningapps.org/

LimeSurvey (Open Source et nécessite PHP) : http://www.limesurvey.org/fr/

One Extra Lap : http://oneextralap.com/

Kahoot : https://getkahoot.com/

Qedoc : http://www.qedoc.org/en/index.php?title=Main_Page

Questar : http://questar.free.fr/

Quizlet : http://quizlet.com/

QuizPoint : http://www.learningware.com/quizpoint.html

QuizPress [payant] : http://www.solrobots.com/quizpress/index.html

QuizRevolution : http://www.quizrevolution.com/

QwickSurveys : http://kwiksurveys.com/

SmarterMarks : http://smartermarks.com/home

TypeForm : http://www.typeform.com/

Quelques sites qui répertorient et ou recensent des outils de construction de QCM :

http://www.quiz-creator.com/blog/2009/09/free-online-quiz-creator-tools-create-online-quizzes/

http://academics.smcvt.edu/cbauer-ramazani/Links/authoring_tools.htm

http://uepd.quizport.com/course/view.php?id=19&topic=5

http://wiki.cyclotronp.net/doku.php/education/exerciseurs/liste_logiciels

http://www.sequane.com/

http://c.deruy.ouvaton.org/telecharger/logiciels-auteurs-libres-gratuits.pdf

http://www.freetech4teachers.com/2012/06/12-tools-for-quickly-gathering-informal.html

http://www.educatorstechnology.com/2014/11/over-40-web-tools-to-create-quizzes-and.html

Un site répertorie quelques logiciels permettant de gérer une banque d'items QCM :

http://fr.wikipedia.org/wiki/Logiciel_de_questionnaire_%C3%A0_choix_multiples

Voici enfin des logiciels permettant le calcul d'analyse d'items :

http://en.wikipedia.org/wiki/Psychometric_software

http://www.rasch.org/software.htm

http://www.assess.com/xcart/product.php?productid=407

http://www.ssicentral.com/irt/index.html

http://languagetesting.info/statistics/excel.html

jMetrik : http://www.itemanalysis.com/

3e Partie : Description du cours MARDIF : Objectifs et organisation

La troisième et dernière partie de ce cours va concerner uniquement les étudiants du cours (MARDIF-CNED) et décrit ses objectifs et son organisation.

12. Objectifs du cours et aspects organisationnels

Objectifs

Ce cours est conçu pour être suivi deux années successives. La première année, les étudiants qui l'auront suivi seront capables :

- de sélectionner l'activité d'apprentissage en lien avec leurs principes pédagogiques ;
- de construire des questions de QCM liées à un cours (également conçu pour l'occasion) qui soient en lien avec les activités d'apprentissage visées dans ce cours ;
- de composer un QCM, dans un contenu donné, qui ne contienne pas trop de biais ;
- de réaliser un bref cours et des QCM assortis avec le logiciel *eXe Learning*.

La deuxième année, les étudiants seront capables :

- d'organiser une passation, sur une vingtaines de personnes, du cours-QCM (niveau et âge approprié à son contenu) conçu la première année ;
- de recueillir et d'analyser les réponses, notamment en termes de difficulté et discrimination ;
- de tirer quelques enseignements à partir de l'analyse de ces réponses (items à garder, à rejeter) ;
- éventuellement (selon leur niveau de maîtrise de l'informatique), de convertir tout ou partie de leur cours sur *Oppia*, un nouveau système de conception de parcours d'apprentissage interactifs.

Chronologie des tâches du séminaire TICE

L'organisation générale du séminaire peut se représenter comme suit dans le Tableau 7. Le détail des tâches pour chaque promotion suit. Rappelons que ce séminaire est conçu pour être passé en 2 ans. La première année les étudiants suivent les activités de la 2e colonne et la deuxième année celles de la 3e.

Tableau 7 – Chronologie des tâches du séminaire TICE, avril 2015

Dates(-Butoir)	Promotion Jean-Bourjade	Activités Promotion Henri-Wallon
1er avril 2015	Démarrage du séminaire	Démarrage du séminaire
Questions formulées le plus tôt possible et avant le 14 avril 2015 à 12:00 CEST	**Activité 1** : Lecture du cours (1re et 2e partie, sauf section 10), installation du logiciel *eXe Learning* (voir Section 9) et formulation de questions	**Activité 1** : Lecture du cours (Section 10), organisation de la passation et formulation de questions
À démarrer au plus tôt	**Activité 2** : Construction et dépôt du cours-QCM créé avec le logiciel *eXe Learning*, dans l'espace de dépôt correspondant (Jean-Bourjade)	**Activité 2** : Réalisation de la passation et analyse des résultats. Rédaction d'un bref rapport. Dépôt du rapport dans l'espace de dépôt correspondant (Henri-Wallon)
30 avril 2015	Clôture du séminaire	Clôture du séminaire

Détail des activités par promotion

A. 1re année (promotion Jean-Bourjade)

Sont décrites ici les tâches des étudiants qui débutent ce séminaire TICE, et qui n'ont donc pas encore eu à réaliser de cours-QCM avec *eXe Learning* l'année précédente.

Activité 1 (voir Tableau 7 pour la date-butoir)

- Lecture extensive et attentive du cours et travail sur les TP (lecture des 1re et 2e parties du cours, hormis la section 10, et travail sur au moins 3 TP, au choix, qui sont des tâches énoncées à la fin de chaque section). Noter que les réponses aux TP sont à garder pour soi, éventuellement à discuter dans le forum, mais pas à livrer à l'enseignant.
- Lecture additionnelle d'au moins 1 autre article conseillé dans la rubrique « voir plus loin » de chaque section.
- Répondre aux questions suivantes de manière argumentée, en une dizaine de lignes par question. Ces questions sont conçues pour vérifier la compréhension des étudiants des tâches demandées et du contenu du cours :
    - Quelles sont les trois principales notions exposées et discutées dans le cours, ainsi que les questions que je me pose à leur propos ? (p. ex., concernant leur possible application à mes propres recherches).
    - Quels sont les trois principaux points du cours qu'il me reste à comprendre ou à approfondir ? Y a-t-il des points particuliers que je n'ai pas compris et sur lesquels j'aimerais avoir des précisions ?
    - Questions techniques. Suis-je arrivé-e à installer le logiciel *eXe Learning* et ai-je compris son fonctionnement dans les grandes lignes ? Quelles sont les questions que je me pose encore, suite à la lecture de la documentation ?
- Dépôt des réponses au questionnaire du cours dans le forum de discussion du séminaire. Retour de l'enseignant sur chacune avant la réunion (lorsque possible, d'où l'intérêt de les poster assez tôt). Le contenu des questions pourra alimenter la discussion de la classe virtuelle.
- Installation du logiciel *eXe Learning* et première manipulation à des fins d'exploration de ses fonctionnalités. L'essentiel des informations à propos du logiciel est mentionné Section 9 (notamment ses modes d'emploi). *Il est important de signaler par courriel à l'enseignant, le plus rapidement possible, tout problème dans l'installation ou l'utilisation du logiciel.* Les étudiants les plus avancés pourront envisager d'utiliser *Oppia*, mais uniquement après avoir réussi à répliquer sans problèmes l'Exploration décrite dans la section 9.
- Pour information, le cours conçu par une étudiante de la promotion MARDIF 2010–11 peut être un bon exemple à suivre (http://perso.numericable.com/annick.pradeau/ViolenceConjugale/index.html). D'autres exemples sont sur le site de l'atelier (Espace : Cours-QCM réalisés l'année précédente).

Activité 2 (voir Tableau 7 pour la date-butoir)

Note : dans la mesure du possible, il est conseillé de *ne pas* attendre la fin de l'activité 1 pour démarrer l'activité 2. Les deux activités peuvent se mener en parallèle.

- Réalisation du cours (présentant un contenu *et* un QCM). Voir Section 9 pour des indications sur ce point. Il est important de noter que votre cours devra à la fois présenter

un contenu de cours (à lire) et un QCM (à compléter) qui aura pour but de vérifier la bonne compréhension du cours.

Sauf avis contraire de l'enseignant, les différents cours-QCM produits sur *eXe Learning* seront installés sur un site public du MARDIF et pourront être librement consultés, à la clôture du séminaire, soit courant juin 2015. À cette fin, chaque étudiant, une fois son cours entièrement conçu sur *eXe Learning*, devra réaliser la procédure suivante :

- Dépôt du cours et du formulaire sur l'espace de dépôt du site MARDIF (correspondant à la promotion de l'étudiant) :

1. Le cours-QCM étant ouvert, exécuter le menu `File>Export>Web Site>Zip File`. Cela enregistre une version compressée du site du cours, dans le format XHTML.
2. Déposer sur l'espace de dépôt approprié du séminaire (celui de votre promotion) les fichiers suivants :
    a. Le formulaire rempli et signé d'autorisation de publication du cours. Si vous n'acceptez pas que votre cours soit porté sur le site, veuillez prévenir l'enseignant du cours qui vous indiquera la procédure alternative.
    b. Le fichier compressé obtenu par l'étape 1
    c. Le fichier source du cours (obtenu en faisant `File>Save As…`), de suffixe .elp.
3. Envoyer un courriel à l'enseignant pour le prévenir du dépôt.

Les critères d'évaluation suivants seront pris en compte dans l'évaluation du cours.

- qualité du contenu du cours (de fond et de forme) ;
- qualité du lien entre cours et QCM ;
- qualité de rédaction des QCM (conformes aux prescriptions de ce cours) ;
- qualité du feedback (donné en retour, en cas de bonne ou mauvaise réponse).

B. 2e année (promotion Henri-Wallon)

Sont décrites ici les tâches des étudiants qui ont déjà, l'année passée, conçu et réalisé les activités mentionnées ci-dessus. Elles se partagent également en deux activités.

Activité 1 – Prise de connaissance et peaufinage du test (voir Tableau 7 pour la date-butoir)

- lire très attentivement la section 10 de ce document pour comprendre de près de quelle manière on calcule différents indices liés au QCM. Parcourir à nouveau le document à la recherche des principales mises à jour (paragraphes avec une fine bordure noire à droite).
- retrouvez dans vos archives le cours-QCM que vous avez conçu l'année précédente, et relisez-le attentivement, notamment pour vous le remettre en tête. Eventuellement, ajoutez ou supprimez quelques questions pour une meilleure cohérence des différents items. En effet (voir section 10 de ce document), l'utilisation d'indices statistiques suppose que les différents items mesurent la même dimension (connaissance, compétence).
- s'assurer que le nombre de questions de type QCM (c.-a.-d., une question-amorce, un choix de réponses,) est au moins égal à 6. S'assurer de plus que chaque item comporte le même nombre de choix, pour rendre l'analyse plus aisée. Cela vous obligera peut-être à rédiger quelques nouveaux items ou affiner les réponses ;

- trouver 20 participants volontaires à la passation du cours-QCM (en présence ou bien à distance, p. ex., autres étudiants du MARDIF, élèves) ;
    - dans le cas où vous démarchez auprès des étudiants du MARDIF, il sera donc nécessaire de placer la version éventuellement modifiée de votre QCM sur le site de l'atelier TICE. Veuillez svp contacter conjointement pour cela par courriel Dominique Wormser dominique.wormser@univ-rouen.fr et Stéphanie Dal Piva stephanie.dal-piva1@univ-rouen.fr, et ce le plus tôt possible (avant le 15 avril 2015) en lui envoyant votre site en .zip sans passer par l'enseignant, qui ne revalide donc pas le cours-QCM. Les sites seront placés dans un espace *ad hoc*, nommé « appel à sondage ». La manière dont les étudiants vont vous informer de leurs résultats (puisqu'*eXe Learning* n'a pas de procédure pour les collecter) est mentionnée dans l'item de niveau supérieur suivant.
    - Pensez à désactiver les rétroactions de votre QCM, de manière à ne récupérer que les premières réactions des participants et éviter de possibles tricheries.
    - Pour une passation avec d'autres participants (élèves, autres étudiants), ce sera à vous, soit de publier votre cours-QCM sur un site, soit de l'exécuter sur des ordinateurs, en local
- penser à un système de collecte des données centralisé (comme les questionnaires *Google Forms*), *ce qui est optionnel* et dépend du niveau de connaissances de l'étudiant, même si faire un questionnaire sur *Google Form* est aisé et qu'il existe sur Internet de nombreux tutoriels d'aide à leur construction. À défaut, vous demanderez aux participants (s'ils sont étudiants du MARDIF) de noter et déposer leurs résultats sur le thème du forum correspondant (ou, s'ils ne sont pas étudiants du MARDIF, par d'autres moyens de votre choix, p. ex. par courriel).
- Répondre aux questions suivantes de manière argumentée dans le forum du séminaire, en une dizaine de lignes par question. Ces questions sont conçues pour vérifier la compréhension des étudiants des tâches demandées et du contenu du cours.
    - Quelles sont les trois principales notions exposées et discutées dans la section 10 du cours, ainsi que les questions que je me pose à leur propos ? (p. ex., concernant leur possible application à mes propres recherches).
    - Quels sont les trois principaux points de la section 10 du cours qu'il me reste à comprendre ou à approfondir ? Y a-t-il des points particuliers que je n'ai pas compris et sur lesquels j'aimerais avoir des précisions ?
    - Questions techniques. Ai-je bien réussi à récupérer mon QCM avec *eXe Learning* et est-ce que j'ai pu inviter une vingtaine de participants à le passer ? Est-ce que je suis également au clair avec la manière dont les réponses au QCM vont me parvenir, et comment vais-je les analyser ?
- Le cas échéant, si votre niveau en informatique est suffisant, utilisez *Oppia* pour convertir une partie du cours.

Activité 2 – Passation, collecte et analyse des données (voir Tableau 7 **pour la date-butoir)**

Note : dans la mesure du possible, il est conseillé de ne pas attendre la fin de l'activité 1 pour démarrer l'activité 2. Les deux activités peuvent se mener en parallèle (notamment la passation du questionnaire, qui peut prendre du temps). L'activité 2 est composée des sous-tâches suivantes :

- faire passer le questionnaire aux 20 participants ;
- récupérer les données de résultats et les placer dans un fichier tableur à des fins d'analyse (le fichier tableur est disponible sur le site de l'atelier et également à http://webu2.upmf-grenoble.fr/sciedu/pdessus/cours/tabQCM.xls). Le cas échéant, il faudra penser à augmenter les lignes et colonnes pour faire correspondre le tableau avec le nombre de vos participants, des items, et du nombre de réponses des items.
- récupérer calculer les différents indices (*P* ou *B* et *ID*) et analyser les items comme indiqué dans la section 10.
- considérer la suppression ou la reformulation de certains items, selon les résultats de l'analyse.
- rédiger un bref document (4 p. maximum) qui sera ensuite mis en format PDF et déposé dans l'espace de dépôt de votre promotion, de structure comme suit :
    - Titre (indiquant le thème du QCM) ; nom et prénom de l'étudiant ; promotion.
    - Bref descriptif du nombre de participants de l'étude, du contexte de passation (présence, distance, à tour de rôle, en parallèle, etc.), de leur niveau scolaire si passation faite en contexte scolaire. Il sera également nécessaire de documenter *a priori* leur niveau (ont-ils un bon ou faible niveau de connaissances du domaine du QCM ?)
    - Mention des indices (*P* ou *B* et *ID*), commentaires.
    - Graphique (voir Figure 3 section 10) indiquant la répartition du choix des réponses des élèves. Il faudra veiller à anonymer les participants du test.
    - Brève analyse, item par item, avec justification des items éventuellement rejetés et proposition d'une nouvelle reformulation de ces derniers (items ou réponses).
    - Copie du tableau entier de résultats.

13. Autorisation de diffusion de mon cours

Document à remplir (mentionner votre nom et prénom dans les ______*, dater et signer, et renvoyer par courriel en format PDF à l'enseignant en même temps que le cours.*

Autorisation entre

_________________________, étudiant-e au MARDIF, contractant-e

et le MARDIF, CNED, représenté par son responsable,

M. Jean-Luc Rinaudo, éditeur,

Je soussigné, le/la contractant-e ____________________, autorise gratuitement [1] / n'autorise pas[1], par la présente, la diffusion du cours *eXe Learning* que j'ai conçu et réalisé dans le cadre de l'atelier TICE du MARDIF, d'avril 2015, sous la forme d'un site internet placé sous la responsabilité éditoriale du CNED. L'enseignant du cours, M. Philippe Dessus, se réserve toutefois la possibilité de ne pas faire publier le cours.

Je suis informé-e que ce cours sera placé selon la licence *Creative Commons* BY-NC-SA, dont les conditions sont disponibles à http://creativecommons.org/licenses/by-nc-sa/3.0/. Je déclare avoir pris connaissance de cette licence et suis en plein accord avec ses conditions.

Je suis pleinement informé-e que le contenu de ce cours sera diffusé par le site intranet *Blackboard* géré par le CNED, MARDIF pendant une durée de 1 an, en libre accès pour tous les étudiants inscrits au MARDIF.

Je certifie également que je suis l'auteur de ce cours, et que aucun élément de ce dernier n'a un droit d'auteur possédé par une autre personne. Si j'ai inséré des images dans mon site, je certifie qu'elles sont libres de droits et peuvent être publiées sur des sites non commerciaux.

Le CNED-MARDIF et l'éditeur ne pourront être tenus responsables dans toute réclamation d'un tiers pour une question de propriété intellectuelle de toute partie du contenu de mon cours.

[1] Barrer la mention inutile.

Nom, prénom, date et signature

14.Références

Andrade, J., Ares, J., Garcia, R., Rodriguez, S., Seoane, M., & Suarez, S. (2008). Guidelines for the development of e-learning systems by means of proactive questions. *Computers & Education, 51*(4), 1510–1522.

Audet, L. (2011). Les pratiques et défis de l'évaluation en ligne. Montréal: REFAD.

Bangert-Drowns, R. L., Kulik, C., Kulik, J., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*(2), 213–238.

Bar-Hillel, M., Budescu, D., & Attali, Y. (2005). Scoring and keying multiple choice tests: A case study in irrationality. *Mind & Society, 4*, 3–12. doi: 10.1007/s11299-005-0001-z

Basque, J., & Lundgren-Cayrol, K. (2002). Une typologie des typologies des applications des TIC en éducation. *Sciences et Techniques Educatives, 9*(3-4).

Bereiter, C. (2002). *Education and mind in the knowledge age*. Mahwah: Erlbaum.

Bernier, J.-J. (1985). *Théorie des tests : Principes et techniques de base* (2e ed.). Chicoutimi: Gaëtan Morin.

Bjork, E. L., Little, J. L., & Storm, B. C. (2014). Multiple-choice testing as a desirable difficulty in the classroom. *Journal of Applied Research in Memory and Cognition, 3*, 165–170. doi: 10.1016/j.jarmac.2014.03.002

Bleske-Rechek, A., Zeug, N., & Webb, R. M. (2007). Discrepant performance on multiple-choice and short answer assessments and the relation of performance to scholastic aptitude. *Assessment & Evaluation in Higher Education, 32*(2), 89–105.

Bloom, B. S. (1988). Le problème des deux sigmas. In M. Hüberman (Ed.), *Assurer la réussite des apprentissages scolaires ?* (pp. 45–83). Neuchâtel: Delachaux et Niestlé.

Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1969). *Taxonomie des objectifs pédagogiques* (M. Lavallée, Trans. Vol. 1 : Domaine cognitif). Montréal: Education Nouvelle.

Bodner, G. M. (1980). Statistical analysis of multiple-choice exams. *Journal of Chemical Education, 57*(3), 188–190.

Bransford, J., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience and school*. Washington: NAP.

Bravard, S. (2005). Usages pédagogiques des QCM : Un guide pour la mise en place d'un questionnaire à choix multiple. Poitiers: Université de Poitier, URF Letttres et langues, mémoire de Master ingénierie des médias pour l'éducation non publié.

Burton, R. F. (2005). Multiple-choice and true/false tests: myths and misapprehensions. *Assessment & Evaluation in Higher Education, 30*(1), 65–72. doi: 10.1080/0260293042003243904

Butler, A. C., & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition, 36*(3), 604–616.

Cheung, D., & Bucat, R. (2002). *How can we construct good multiple-choice items?* Paper presented at the Science and Technology Education Conference, Hong Kong.

Clarke, M. M., Madaus, G. F., Horn, C. L., & Ramos, M. A. (2000). Retrospective on educational testing and assessment in the 20th century. *Journal of Curriculum studies, 32*(2), 159–181.

Crisp, G. T., & Palmer, E. J. (2007). Engaging academics with a simplified analysis of their Multiple-Choice Question (MCQ) assesment results. *Journal of University Teaching and Learning Practice, 4*(2), 88–106.

Davidson, C. N. (2011). *Now you see it. How the brain science of attention will transform the way we live, work, and learn*. New York: Vicking.

de Vries, E. (2001). Les logiciels d'apprentissage : panoplie ou éventail ? *Revue Française de Pédagogie, 137*, 105–116.

Demaizière, F. (1986). *Enseignement assisté par ordinateur*. Paris: Ophrys.

Depover, C. (1987). *L'ordinateur media d'enseignement*. Bruxelles: De Boeck.

Dessus, P. (2000). La planification de séquences d'enseignement, objet de description ou de prescription ? *Revue Française de Pédagogie, 133*, 101–116.

Dessus, P., & Lemaire, B. (2004). Assistance informatique à la correction de copies. In E. Gentaz & P. Dessus (Eds.), *Comprendre les apprentissages : sciences cognitives et éducation* (pp. 205–220). Paris: Dunod.

Enard, C. (1970). Le guidage dans l'apprentissage. In J. Leplat, C. Enard & A. Weill-Fassina (Eds.), *La formation par l'apprentissage* (pp. 113–156). Paris: P.U.F.

Erstad, O. (2008). Changing assessment practices and the role of it. In J. Voogt & G. Knezek (Eds.), *International handbook of information technology in primary and secondary education* (pp. 181–194). New York: Springer.

Fellenz, M. R. (2004). Using assessment to support higher level learning: the multiple choice item development assignment. *Assessment & Evaluation in Higher Education, 29*(6), 703–719. doi: 10.1080/0260293042000227245

Francis, A. L., & Barnett, J. (2012). The effect and implications of a "self-correcting" assessment. *Teaching of Psychology, 39*(1), 38–41.

Georges, F., & Pansu, P. (2011). Les feedbacks à l'école : un gage de régulation des comportements scolaires. *Revue Français de Pédagogie, 176*, 101–124.

Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3th ed.). Mahwah: Erlbaum.

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education, 15*(3), 309–334.

Hanna, G. S., & Oaster, T. R. (1978). Toward a unified theory of context dependence. *Reading Research Quarterly, 14*(2), 226–243.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Johnstone, A. (2003). Effective practice in objective assessment. The skills of fixed response testing. Hull (UK): LTSN Physical Sciences.

Kalz, M. (2009). *Placement support for learners in learning networks*. Heerlen: OUNL, unpublished PhD Thesis.

Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research, 47*(1), 211–232.

Laveault, D., & Grégoire, J. (2002). *Introduction aux théories des tests en psychologie et en sciences de l'éducation* (2e ed.). Bruxelles: De Boeck.

Le Clainche, E. (2010). The eLearning XHTML editor: Guide [de] l'utilisateur. Rennes: ISFEC de Bretagne.

Leclercq, D. (2006). L'évolution des QCM. In G. Figari & L. Mottier-Lopez (Eds.), *Recherches sur l'évaluation en éducation* (pp. 139–146). Paris: L'Harmattan.

Lemercier, C., Tricot, A., Chênerie, I., Marty Dessus, D., Morancho, F., & Sokoloff, J. (2001). Quels apprentissages sont-ils possibles avec des exerciseurs multimédia en classe ? Réflexions théoriques et compte rendu d'une expérience. Paris: Rapport de recherche, contribution au rapport du Programme de Numérisation de l'Enseignement et de la Recherche « Usages éducatifs des exerciseurs ».

Luengo, V. (2009). Les rétroactions épistémiques dans les Environnements Informatiques pour l'Apprentissage Humain. . Grenoble: Université Joseph-Fourier, dossier pour l'habilitation à diriger des recherches.

Madaus, G. F., & O'Dwyer, L. M. (1999). A short history of performance assessment: Lessons learned. *Phi Delta Kappan, 80*(9), 688–695.

Marsh, E. J., Roediger, H. L., 3rd, Bjork, R. A., & Bjork, E. L. (2007). The memorial consequences of multiple-choice testing. *Psychonomic Bulletin & Review, 14*(2), 194–199.

Mayer, R. E., & Colvin Clark, R. (2010). Instructional strategies for receptive learning environments. In K. H. Silber & W. R. Foshay (Eds.), *Handbook of improving performance in the workplace: Vol. 1, Instructional design and training delivery* (pp. 298–328). San Francisco: Pfeiffer.

McAlpine, M. (2002b). Principles of assessment. Loughborough University: CAA Centre.

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology, 19*(4/5), 494–513. doi: 10.1080/09541440701326154

McDaniel, M. A., Wildman, K. M., & Anderson, J. L. (2012). Using quizzes to enhance summative-assessment performance in a web-based class: An experimental study. *Journal of Applied Research in Memory and Cognition, 1*, 18–26. doi: 10.1016/j.jarmac.2011.10.001

Means, B., Toyama, Y., Murphy, R., & Bakia, M. (2010). Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies. Washington: U.S. department of education.

Monaham, T. (1998). The rise of standardized educational testing in the U.S.: A bibliographic overview. Albany: Report of the Rensselaer Polytechnic Institute.

Montepare, J. M. (2005). A self-correcting approach to multiple choice tests. Retrieved from http://www.psychologicalscience.org/observer/getArticle.cfm?id=1867

Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multi-media learning. In H. Niegeman, R. Brünken & D. Leutner (Eds.), *Instructional design for multimedia learning* (pp. 181–195). Münster: Waxmann.

Osman, R. M. (2011). Mesure et évaluation en éducation. s.l.: Cours de l'université virtuelle africaine.

Osterlind, S. J. (1998). *Constructing test items* (2th ed.). New York: Springer.

Reif, F. (2008). *Applying cognitive science to education*. Cambridge: MIT Press.

Rodriguez, M. C. (1997). *The art and science of item-writing: A meta-analysis of multiple-choice item format effects* Paper presented at the American Educational Research Association, Chicago.

Roediger, H. L., 3rd, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(1), 20–27. doi: 10.1016/j.tics.2010.09.003

Roid, G., & Haladyna, T. M. (1980). The emergence of an item-writing technology. *Review of Educational Research, 50*(2), 293–314.

Sadler, P. M. (1999). The relevance of multiple-choice tests in assessing science understanding. In J. J. Mintzes, J. H. Wandersee & J. D. Novak (Eds.), *Assessing science understanding: A human constructivist view* (pp. 249–278). San Diego: Academic Press.

Seyve, D., & Grépilloux, J. (2004). TICE et QCM. Grenoble: GRECO, les dossiers thématiques du projet GRECO.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189.

Touchie, C. (s.d.). Guidelines for the development of multiple-choice questions. Ottawa: Medical Council of Canada.

Towns, M. H. (2014). Guide to developing high-quality, reliable, and valid multiple-choice assessments. *Journal of Chemical Education, 91*(9), 1426–1431. doi: 10.1021/ed500076x

Tractenberg, R. E., Gushta, M. M., Mulroney, S. E., & Weissinger, P. A. (2013). Multiple choice questions can be designed or revised to challenge learners' critical thinking. *Advances in Health Sciences Education, 18*(5), 945–961.

Trausan-Matu, S., Dessus, P., Lemaire, B., Mandin, S., Villiot-Leclercq, E., Rebedea, T., . . . Zampa, V. (2008). Deliverable D5.1 LTfLL – Support and feedback design. Heerlen: OUNL, Research report of the LTfLL Project.

Williams, J. B. (2006). Assertion-reason multiple-choice testing as a tool for deep learning: a qualitative analysis. *Assessment & Evaluation in Higher Education, 31*(3), 287–301.